# INVESTIGATION OF THE INTERACTIONS IN RNA RECOGNITION MOTIF-RNA COMPLEXES: SMALL MOLECULE INHIBITORS AND KINETICS OF DISSOCIATION

BY

JUNG-UN BAEK

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Anne M. Baranger, Chair
Professor John A. Katzenellenbogen
Professor Paul J. Hergenrother
Professor Huimin Zhao

## ABSTRACT

The RNA recognition motif (RRM) is the most abundant RNA binding domain that is found in all organisms. RRM-containing proteins participate in most steps of gene expression, including translation, splicing, modification and transport of RNA. This dissertation aims to help understand the interactions of the RNA recognition motif and RNA by developing a small molecule modulator of the interaction and analyzing the kinetics of dissociation. The first chapter gives an introduction to the function and structural characteristics of RNA binding proteins, RNA recognition motifs and two RRM proteins, U1A and Sex lethal protein. Chapter 2 describes the identification and analysis of three small molecules that disrupt two different RRM-RNA complexes, Sex lethal protein-*tra* RNA and U1A-SL2 RNA. The research discussed in chapter 3 focus on the role of positively charged residues in the U1A protein and SL2 RNA complex dissociation process. Analysis of kinetics data obtained by temperature jump and stopped-flow experiments showed that the location of the electrostatic interaction controls the rate of different steps in the complex dissociation pathway. Chapter 4 is a description of a simple and rapid method to detect RNA splice variants using biarsenical dyes and split tetracysteine moieties, which may accelerate biochemical studies of alternative splicing and identification of factors that modulate RNA splicing.

## ACKNOWLEDGEMENTS

This work would not have been possible without the support of many people: First and foremost, my advisor, *Dr. Anne M. Baranger*, who offered invaluable assistance, support and guidance. She brought out the best ideas in me and encouraged me to hang on when I doubted myself. My committee members, *Professors John A. Katzenellenbogen, Paul J. Hergenrother, Huimin Zhao* and my Chemical Biology Interface advisor *Professor Wilfred A. van der Donk* for their advice and support. My research collaborators, *Professor Martin Gruebele* and *Irisbel Guzman*, who gave me excellent feedback and ideas for the project and motivated me to do more productive research.

I am thankful to the former and current members of the Baranger Group: *Divina*, *Doug*, *Yan*, *Sreeni*, *Stacie*, *Adam*, and *Yuan* who were wonderful friends and answered all my research related questions.

I would like to acknowledge the greatest blessing that I ever have – *my parents* for their unconditional love and untiring support. Special thanks to *my husband, Andy* who helped me with research and made me smile whenever I was depressed.

## TABLE OF CONTENTS

# CHAPTER 1

# RNA Binding Proteins and RNA Recognition Motif

## *1.1 RNA binding proteins (RBPs)*

RNA is a highly versatile molecule that plays a fundamental role in all living organisms. Its functions range from being a messenger of genetic information encoded by DNA to the regulations of essential cellular processes.[1-4] In most cases, these functions are carried out in close association with RNA-binding proteins.[4-7] RNA binding proteins (RBPs) bind to RNA as soon as RNA is transcribed and regulate RNA metabolism including pre-mRNA splicing, polyadenylation, mRNA export, nonsense mediated decay and protein translation.[5,6] Several RBPs may interact with RNA to form functional units called ribonueloprotein (RNPs) or may individually bind to *cis*-acting elements of the RNA. There are several hundred RBPs in humans and this number is likely to increase with advanced genetic, biochemical and computational methods.[8] The most prominent structural feature of RBPs is that they contain RNA binding domains as well as auxiliary domains that mediate protein-protein interactions or catalytic activity.

### *1.1.1 Functional diversity of RNA binding proteins*

From genomics to proteomics, RNA binding proteins (RBPs) play an essential role in gene expression and gene regulation at various levels (Figure 1.1).[6,9,10] Many RBPs carry out a wide variety of functions by forming ribonucleoprotein complexes (RNPs). RNPs are highly dynamic as different RBPs bind and dissociate at different stages during post-transcriptional processes. Depending on where and when the RBPs associate with the RNA transcript, they are

classified differently. The primary protein-coding transcripts that are produced by polymerase II are termed pre-mRNA or heterogeneous nuclear RNAs (hnRNPs). Therefore proteins that bind to pre-mRNA in the nucleus are termed hnRNP proteins.[5] On the other hand, proteins that bind to mature RNA messages in the cytoplasm are termed mRNP proteins. However this nomenclature is not necessarily accurate since some RBPs are bound before and after RNA splicing and can shuttle between nucleus and cytoplasm. hnRNP and mRNP proteins participate in transcriptional regulation, biogenesis of telomere, immunoglobin gene recombination, splicing, pre-ribosomal-RNA processing, 3' end maturation, nucleo-cytoplasmic transport of mRNA, cytoplasmic localization, translation, and stability.[5,11]

Non-coding RNAs play a significant role in numerous cellular processes. ncRNAs in complex with RNA binding proteins form various RNPs. These RNPs function on many target RNAs including mRNA, ribosomal RNA, snRNAs, and tRNA.[4] The small nuclear (sn)RNPs and small nucleolar (sno) RNPs are the most studied ncRNPs. snRNPs form the core of spliceosome with the exception of U7 snRNP, which is involved in histone pre-mRNA 3' end formation. snRNPs are composed of U snRNA (U1, U2, U4, $U4_{atac}$, U5, U7, U11 or U12), a heptameric ring of Sm proteins and other proteins specific to each snRNP. snRNP proteins are responsible for the biogenesis, nuclear export, and the activity of the snRNPs.

While there are less than a dozen snRNAs in eukaryotic cells, there are more than 200 unique snoRNAs.[12] snoRNAs guide the modification of nucleotides by base-pairing to the target RNAs.[13] Based on the conserved sequence elements, snoRNAs are divided into two major families, the C/D and H/ACA RNAs. All of the RNAs of each family bind to a distinct set of highly conserved RBPs to form C/D RNPs and H/ACA RNPs. The primary function of C/D RNPs is to methylate the 2'-OH group of target nucleotides, whereas H/ACA RNPs convert

uridine to pseudouridine. Dyskerin directly binds to H/ACA RNA through its RNA binding domain and mediates pseudouridylation of rRNAs and snRNAs. For C/D RNPs, L7Ae binds to the C/D RNA and creates another binding site for the NOP56 protein. NOP56 in turn recruits the catalytic protein fibrollarin, which is responsible for the 2'-O-methylation of nucleotides.[4] Most snoRNPs modify other ncRNAs including rRNA, snRNA, tRNA. However at least one pre-mRNA was found to be targeted by C/D RNP. The brain specific C/D snoRNA, HBII-52 binds to the pre-mRNA of seratonin receptor and regulates alternative splicing.[14] Another interesting target of snoRNP is the telomere. Telomerase RNA is a type of H/ACA RNA which forms a telomerase complex. Telomerase reverse transcriptase (TERT) is the RBP in the complex that extends the ends of chromosomes. Targets of a large number snoRNAs remain to be discovered. Identification of the substrates of snoRNPs might further expand the function of snoRNPs.

RBPs regulate biogenesis, localization, degradation, and the activity of microRNAs, which in turn repress gene expression by interacting through the 3' UTR of mRNAs.[15] Transcribed miRNAs are processed by Drosha and then exported to the cytoplasm by Exportin 5 and RAN-GTP. In the cytoplasm, they are further processed by Dicer and HIV-1 transactivating response RNA binding protein (TRBP). Argonaute protein is recruited by TRBP and the miRNA eventually forms the miRNA-induced silencing complex (miRISC), also called the miRNA ribonucleoprotein complex (miRNP). miRNP consists of miRNA, Ago protein and another RNA binding protein, GW182. This complex is responsible for the translational inhibition of various mRNAs. Other than snRNPs, snoRNPs and miRNPs, there are many ncRNPs that function in diverse biological processes: pre-tRNA maturation by RNase P, transcription elongation by 7SK RNP, protein translocation to the endoplasmic reticulum by SRP RNP etc.[4]

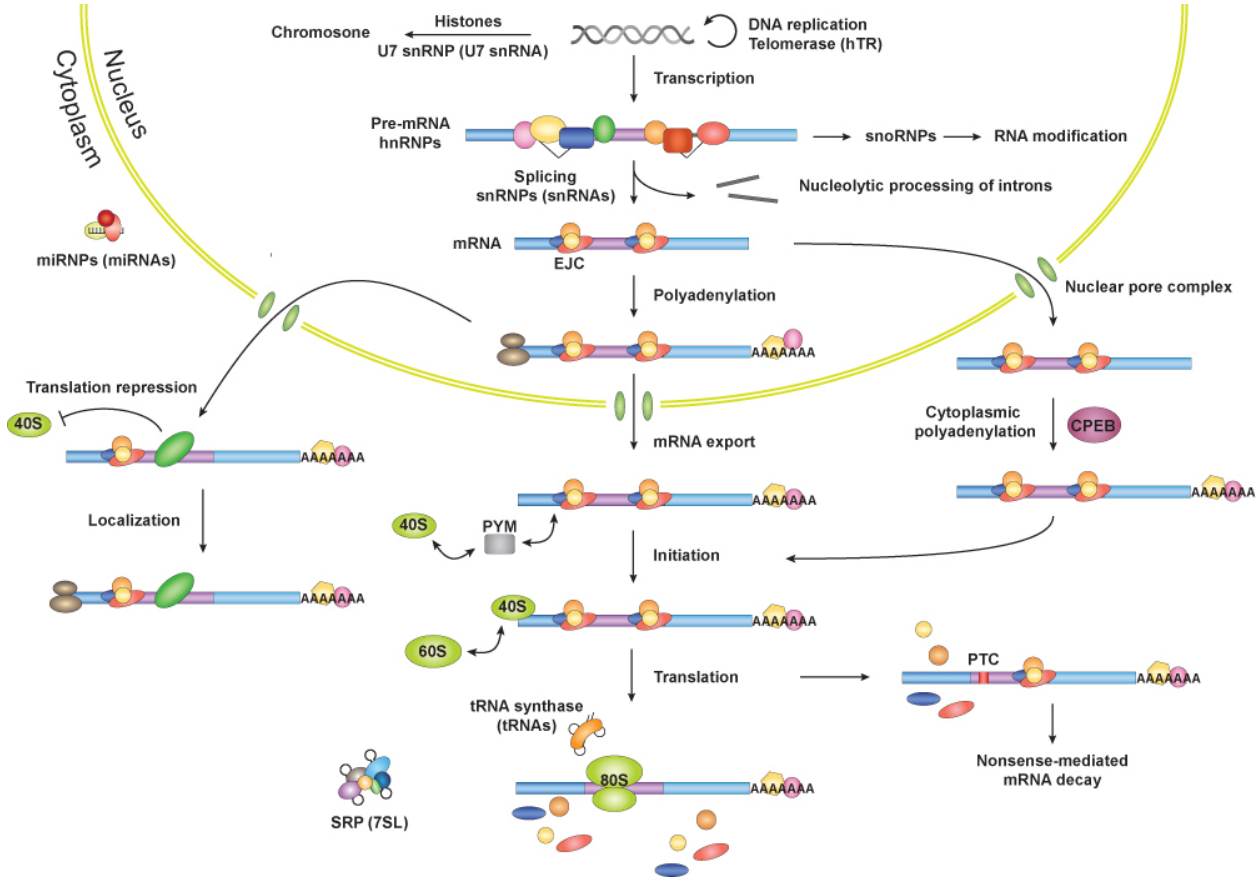



**Figure 1.1** The function of RBPs in multiple cellular processes. Nascent transcripts that are produced by RNA polymerase II are recognized by diverse RBPs such as hnRNP proteins and SR (serine/arginine-rich) proteins. Pre-mRNA splicing is carried out by uridine-rich small nuclear RNPs (U snRNPs) that make up the spliceosome. Spliced mRNAs are associated with RBPs that are organized into a complex near exon-exon junctions (EJC). EJC proteins direct the export, cytoplamic localization or non-sense mediated decay (NMD) of mRNAs. Histone mRNAs are processed by U7 snRNP. Telomerase is a ribonucleoprotein enzyme that replenishes the terminal telomeric repeats of chromosomes. Small nucleolar RNPs (snoRNPs) mediate maturation of ribosomal RNAs. In cytoplasm, ribosomes translate mRNA into proteins. It also functions with the signal recognition particle (SRP) RNP to carry proteins into the endoplasmic reticulum (ER). MicroRNAs interact with RBPs to regulate transcription, mRNA stability and translation.[6]

*1.1.2 RNA binding domains*

RBPs have a modular structure and contain RNA binding domains that mediate RNA recognition (Figures 1.2 and 1.3). Most RBPs consist of multiple copies of RNA binding domains presented in various structural arrangements to specifically bind to diverse RNA substrates. RNA binding domains of the same or different structural types are connected with linkers to form a large binding surface, which allows the protein to recognize a longer stretch of RNA. This modularity also enables the protein to bind to sequences that are separated by an intervening stretch of nucleotides.

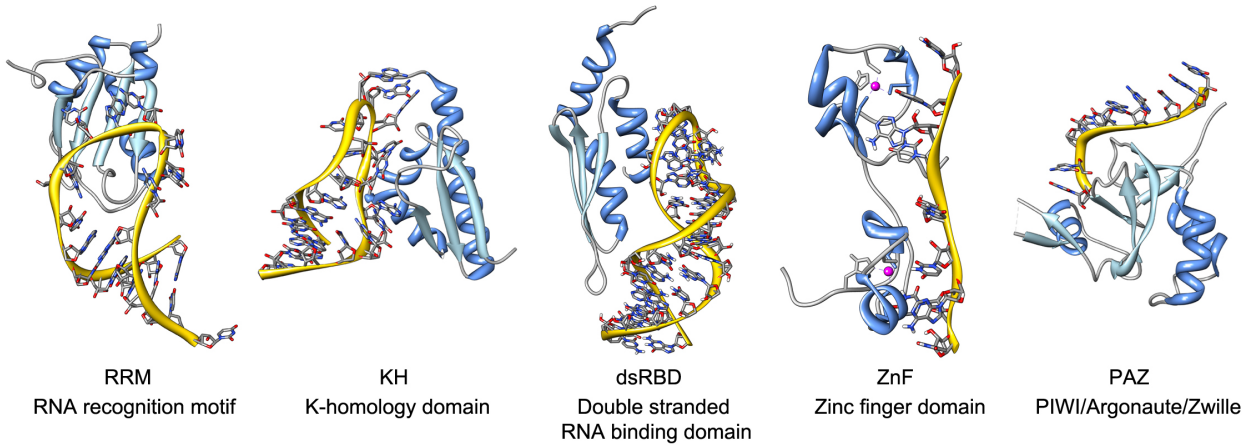



**Figure 1.2** Structure of RNA binding domains in complex with RNAs. RNAs are highlighted in yellow, α-helices are colored blue, and β-sheets are in light blue. pdb codes for each domains are: RRM (1URN), KH (1EC6), dsRBD (1T4L), ZnF (1RGO), PAZ (1SI3). These structures were generated with the program Chimera.[16]

To date, 40 different types of domains that are involved in RNA recognition are identified and the number is constantly increasing.[9] Some well-characterized RNA binding domains include the following: RNA recognition motif (RRM, RBD or RNP), K-homology (KH) domain, double-stranded RNA binding domain (dsRBD), Zinc finger, PIWI/Argonaute/Zwille

(PAZ) domain, and Pumilio domain. The topology, primary RNA recognition mode, and type of protein-RNA interactions of these RNA binding domains are listed in Table 1.1. The RNA recognition motif is the most abundant RNA binding domain and will be discussed in more detail in section 1.3.

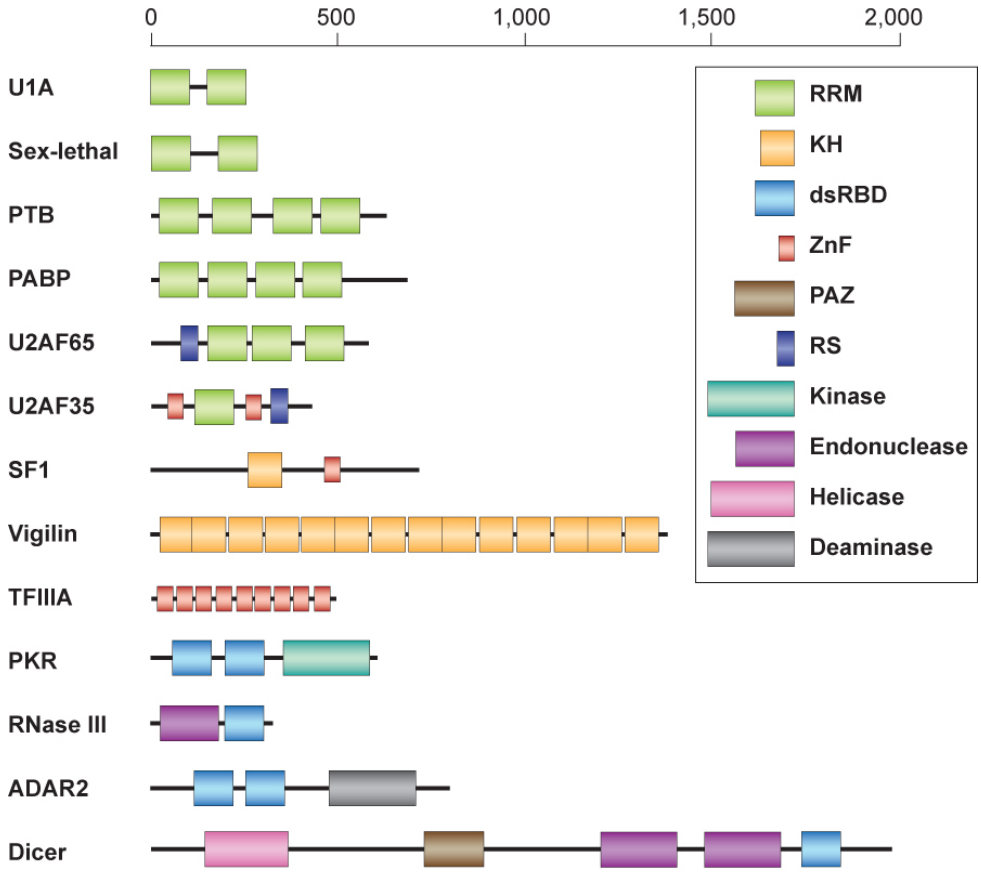



**Figure 1.3** Modular structure of RBPs. RBPs are built from RNA binding motifs. Some RBPs also contain auxiliary motifs, which mediate protein-protein interactions (e.g. RS domain) or catalytic activity (e.g. kinase, endonuclase, helicase and deaminase).[17]

RBPs frequently contain domains that mediate protein-protein interaction or catalytic activity. Proteins that assist spliceosomes in splice site selection such as splicing factors and hnRNPs contain arginine/serine-rich (RS) domains, which participate in protein-protein interaction. Some RBPs contain enzymatic domains, including kinase, helicase, endonuclease,

and deaminase. RNA binding domains locate and in some cases regulate the activity of these catalytic domains. For instance, Drosha and Dicer position the two catalytic RNase III domains using RNA binding domains. In protein kinase PKR and the RNA-editing enzyme ADAR2, RBDs interact with the catalytic domain to maintain the inactive state. Upon RNA recognition the catalytic domain is released and activated.

**Table 1.1** RNA recognition of common RNA binding domains (RBDs).[17]

| Domain | Topology | RNA recognition surface | Protein-RNA interactions |
|---|---|---|---|
| RRM | βαββαβ | Surface of β-sheet | Interacts with 2-8 nt of ssRNA<br>Stacking, electrostatics, H-bond |
| KH Type I | βααββα | Hydrophobic cleft formed by variable loop between β2, β3 and GXXG loop | Interacts with ~4 nt of ssRNA or ssDNA<br>Hydrophobic interactions between non-aromatic residues and the bases<br>H-bond, electrostatics |
| KH Type II | αββααβ | Same as type I, except variable loop is between α2 and β2 | |
| dsRBD | αβββα | Helix α1, N-terminal portion of helix α2, and loop between β1 and β2 | Shape specific recognition of the minor-major-minor groove pattern of dsRNA through contacts to the sugar-phosphate backbone |
| ZnF-CCHH | ββα | Primarily residues in α-helices | Protein side chain contacts to bulged bases in loops and through electrostatic interactions between side chains and the RNA backbone |
| ZnF-CCCH | – | Aromatic side chains form hydrophobic binding pockets for bases that make direct hydrogen bonds to protein backbone | Stacking interactions between aromatic residues and bases create a kink in RNA that allows for the direct recognition of Watson-Crick edges of the bases by the protein backbone |
| PAZ | – | Hydrophobic pocket formed by OB-like β-barrel and small αβ motif | Recognizes ssRNA<br>Stacking interactions and H-bonds |
| Pumilio | Repeats of ααα | Two repeats combine to form binding pocket for individual bases; helix α2 provides specificity-determining residues | Binding pockets for bases provided by stacking interactions<br>Specificity dictated by H-bonds to the Watson-Crick face of a base by two amino acids in helix α2 |

### *1.2 RNA recognition motif (RRM)*

The RNA recognition motif (RRM, also known as RBD or RNP motif) is the most abundant and best characterized RNA binding domain.[18] It is found in all life kingdoms, including prokaryotes and viruses although it is most prevalent in eukaryotes. According to the Pfam database, more than 30,000 sequences have been identified to encode RRM. RRMs have

been implicated in various cellular functions including spliceosome assembly,[19] RNA chaperone activity,[20] and telomere biogenesis.[21] The RRM is composed of approximately 90 amino acids and is structurally characterized by a typical βαββαβ-fold. Structural and biochemical studies have shown that RRM is a simple, but highly versatile, protein motif that can bind to not only RNA but also DNA and proteins, reflecting its functional diversity.

*1.2.1 Primary binding mode of RRM*

The RRM is comprised of about 90 amino acids that fold into an antiparallel β-sheet flanked by two α-helices with a β1α1β2β3α2β4 topology (Figure 1.4). RRMs typically bind to single-stranded RNA in a variety of structural contexts. Recognition of RNA by RRMs relies on a combination of stacking interactions, hydrogen bonds, van der Waals contacts, and salt bridges. Based on the structural analysis of many RRM-nucleic acid complexes, common themes of RRM-nucleic acid interaction have been described. Two examples of RRMs, hnRNPA1 RRM2 and polyadenylate binding protein (PABP), in complex with DNA and RNA, respectively, are shown in Figures 1.4a and b. The beta sheet is the primary surface for single-stranded nucleic acid recognition. The RRM has two consensus sequences, RNP1 and RNP2, in the β3 and β1 strands, respectively (Figure 1.4c). In both structures, three positions of the RNP1 and RNP2 have aromatic residues (position 2 of RNP2, position 3 and 5 of RNP1). Two of the aromatic residues (position 2 of RNP2 and position 5 of RNP1) form stacking interaction with the RNA bases. The third aromatic residue located on β3 (position 3 of RNP1) has different roles in RNA recognition between the two complexes. In hnRNPA1 this residue interacts hydrophobically with sugar rings of the RNA. In PABP, the hydroxyl group of Tyr54 forms electrostatic interactions with the phosphate backbone of A8 and the 2'-OH group of A7 of the RNA. Within

the consensus sequence this variation in the mode of binding is rather common. This trend is also found with the first residue of RNP1. In many RRM proteins, this residue is an arginine or a lysine that contacts an RNA base or a phosphate oxygen of other nucleotides. In hnRNPA1 this residue is an arginine that forms a salt bridge with the phosphate between A203 and G204. However, this residue in PABP is a leucine that forms van der Waals interactions with the A7 of the RNA.

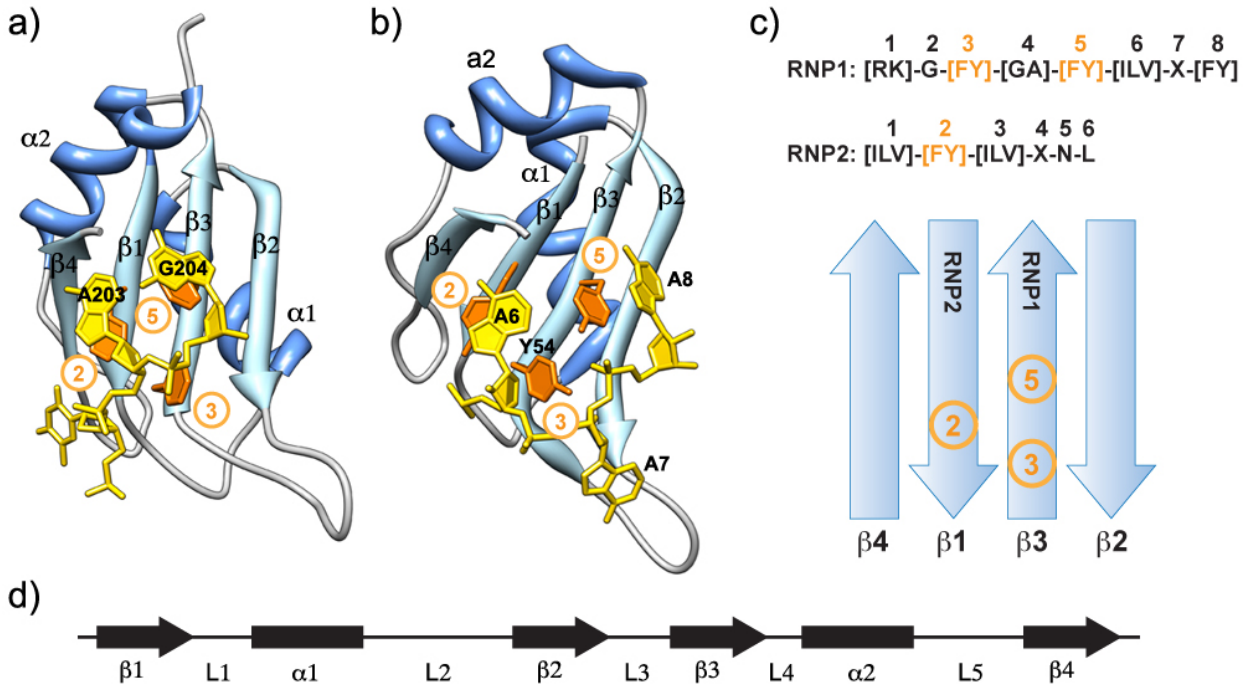



**Figure 1.4** The interaction of the β-sheet surface of the RRM with a single-stranded nucleic acid. a) RRM2 of hnRNP A1 protein bound to telomeric DNA (pdb: 2UP1). b) RRM1 of PABP bound to polyadenylate RNA (pdb: 1CVJ). Nucleotides are shown in yellow, α helices of the protein in blue, β-sheet in light blue, and other regions of the protein in grey. These structures were generated with the program Chimera.[16] c) Consensus sequences of RNP1 and RNP2 and the scheme of the four β-sheets with the conserved residues highlighted in orange. X can be any amino acid.[18] d) A linear scheme of the secondary structure. L is loop.

As exemplified with hnRNP A1 and PABP, most RRMs form at least two stacking interactions with RNA using two of the three conserved aromatic residues. These π-π stacking interactions are critical for the binding affinity of the RRM to the RNA. The substitution of one of these residues for an alanine results in a 100,000 fold decrease in binding affinity in U1A (F56A)[22] and 30,000 fold decrease in binding in Fox-1 (F160A).[23] This basic binding platform common to all RRMs, has a minimal role in sequence-specificity of the RRM. Eight of the 16 possible dinucleotide combinations have already been found to be recognized by proteins containing RRMs, indicated in parenthesis: AA (PABP), AG (hnRNP A1), CG (nucleolin), CA (U1A), GU (Sex lethal), UC (nucleolin), UG (Fox-1) and UU (Sex lethal).[18] Specificity of RNA recognition is provided by other residues in the consensus sequence and non-conserved elements of the RRM, mostly via base specific hydrogen bonds.

*1.2.2 Role of other regions in binding*

Other than the primary binding mode that uses the two aromatic residues for stacking interactions, the binding pattern is significantly different from one RRM to the other.[18,24] This variation in binding mode explains how this structurally conserved motif can bind to different nucleic acid targets. The two external β-strands (β2 & β4), loops, N-terminal region, and C-terminal region may be utilized to different extents for target recognition. For example, in RRM1 of hnRNPA1, the β4 and the β2 each interact with one nucleotide,[25] while most RRMs do not use these β-strands to contact nucleotides.[18]

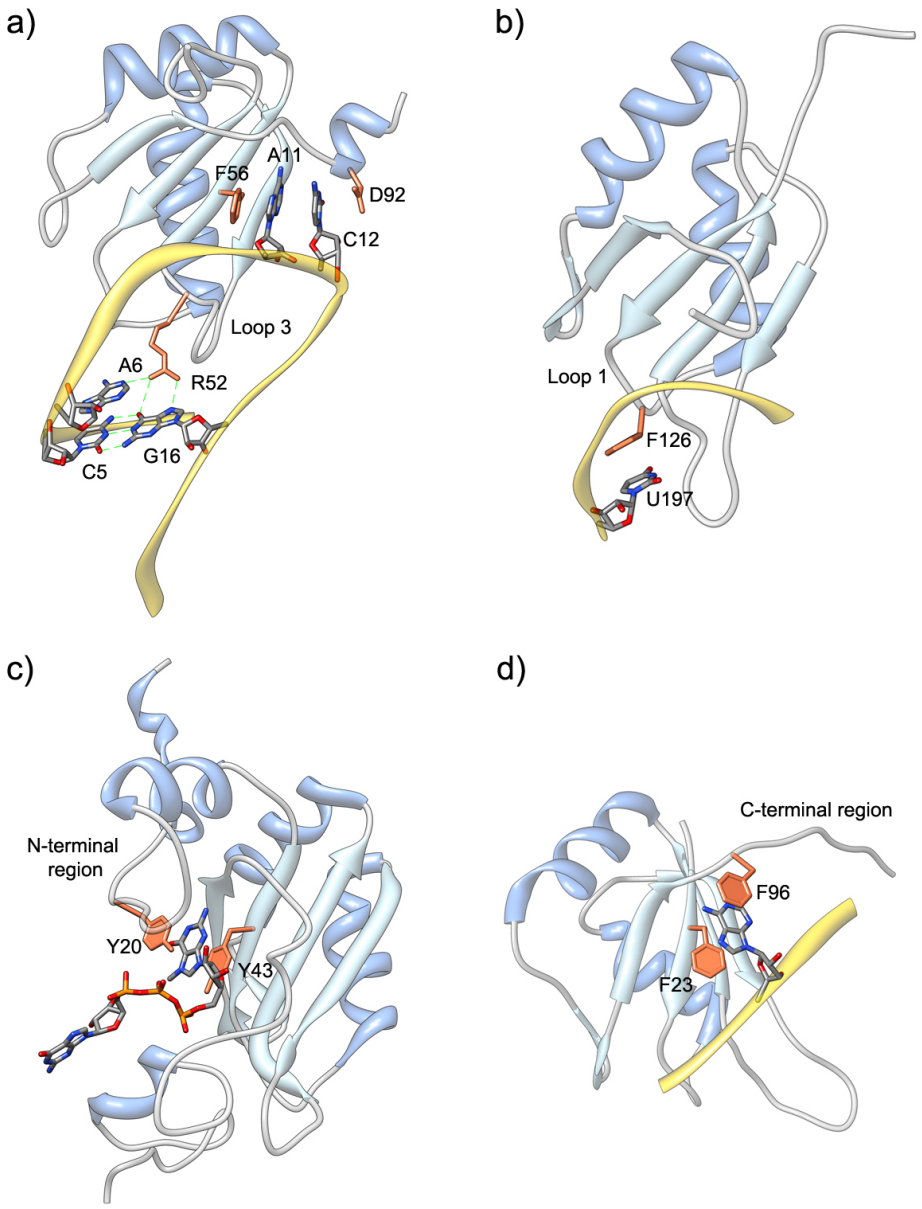



**Figure 1.5** Role of other regions in RNA binding. a) N-terminal RRM of U1A bound to SL2 RNA (pdb:1URN). b) Fox-1 bound to 5'-UGCAUGU-3' RNA (pdb: 2ERR). c) CBP20 bound to 7-methyl guanosine cap ($m^7$GpppG) (pdb:1N52). d) RBD1 of Musashi1 bound to *numb* RNA (pdb: 2RS2). These structures were generated with the program Chimera.[16]

In many cases, loops connecting the β-strand and the α helix are crucial for nucleic acid recognition. For example, loop 3 of the N-terminal RRM of U1A is crucial for the recognition of

stem loop 2 RNA. Several hydrophobic contacts are made to insert the loop into the RNA stem loop and a critical residue, Arg52, forms hydrogen bonds with A6 in the loop of the RNA and G16 of the GC stem (Figure 1.5a). Arg to Gln mutation at this position completely abolishes binding affinity to the RNA.[26] Another RRM protein, Fox-1 has a phenylalanine residue in loop 1 that forms a stacking interaction with the base of U197 (Figure 1.5b).[23] Mutation of this residue to an alanine resulted in a 1,500-fold decrease in binding affinity, which is comparable to the amount of decrease that is found with an aromatic residue of one of the RNP sequences is mutated to an alanine. This type of aromatic residue is found in several other RRMs including REF2-I but is not a general feature of RRMs.

In an extreme case, the β-strands are not used to bind RNA and the binding is primarily accomplished by the loops. This was found in three RRMs of hnRNP F, which lack the conserved aromatic residues in the RNP.[27] These RRMs have been renamed quasi RRMs. Cléry *et al.* have suggested that quasi RRMs may have appeared later in evolution in order to expand the range of RNA target sequences that can be bound by an RRM.[24]

N-terminal and C-terminal regions often interact with the RNA and enhance the RNA-binding affinity. In many RRM-RNA complexes, the base stacking on the aromatic residue at position 2 of RNP2 is sandwiched either by a protein side chain from the N-terminal region or by one from the C-terminal region (Figure 1.5c and d). In some RRMs, the C-terminal region contains a short α helix. This C-terminal helix (helix C) in free U1A protein is contacting the β-strand surface where the RNA binds.[28] In the process of RNA binding, helix C moves away from the β-surface and stays away in the complex. The bound conformation allows the RNA to contact the primary β-surface and form various contacts with the helix C including stacking interaction of Asp92 with C12 (Figure 1.5a).[26] Helix C of U1A is thought to contribute to the

specificity of the RRM through participation in dynamical processes.[29,30] In another protein CstF-64, the helix C unfolds upon binding and does not have any direct contact with the RNA.[31,32]

*1.2.3 Recognition of RNA over DNA*

RRMs usually interact with single stranded RNA, but they are also known to interact with single stranded DNA.[24] Some RRMs have a much higher binding affinity to the single stranded RNA compared to the same sequence of single stranded DNA as found in Sex lethal protein and Polyadenylate binding protein (PABP).[33-35] Other RRMs, such as hnRNP A1, bind specifically and with high affinity to both DNA and RNA of similar sequences.[25] RRMs that specifically recognize single stranded RNA form hydrogen bonds with the 2'-OH group of the sugar ring with the side chain or the amide group of the backbone. These hydrogen bonds are mediated by various parts of RRM including C-terminus, β-strands and the loops. Other than the direct protein-RNA contacts, indirect effects through structural requirements for the RNA may also play an important role in RNA recognition. For example, when *tra* RNA is bound to Sex lethal it forms a kink in the middle, which is mediated by three internucleotide hydrogen bonds between the 2'-OH groups and phosphate groups of the RNA.[36]

*1.2.4 Proteins with multiple RRMs*

Most RRM proteins contain multiple RRMs. While some proteins need just one RRM to bind to their target nucleic acids, other proteins require two or more RRMs to bind to the RNA. Using two or more RRMs, the protein may expand the RNA binding surface to recognize a longer stretch of nucleic acid. Most of the structures solved to date show that RRMs bind to a

stretch of consecutive nucleotides (8-10 nucleotides). However the NMR structure of polypyrimidine tract binding protein (PTB) bound to the polypyrimidine tract of *c-src* and biochemical studies revealed how RRMs may recognize two separated stretches of nucleotides.[37] RRM3 and RRM4 of PTB are tightly bound to each other via their α helices resulting in an anti-parallel orientation of their bound sequences of RNA (Figure 1.6a). The two bound sequences can be separated by more than 15 nucleotides, which form a RNA loop.[37,38] PTB may bind to separate polypyrimidine tracts within a pre-mRNA and loop out the exon or the branch point to repress splicing of the mRNA. U2AF is also thought to bind to two separated polypyrimidine tracts using two RRMs and the flexible interdomain linker.[39] Most proteins that contain multiple RRMs have a flexible linker which often plays an essential role in RNA recognition. The association of multiple RRMs with flexible interdomain linkers provides a range of possible conformations (Figure 1.6). The two RRMs of PTB are bound back-to-back so that the RNA binding surfaces are facing the opposite direction (Figure 1.6a).[37] The two RRMs of polyadenylate binding protein (PABP) provide a continuous binding surface (Figure 1.6b). In Sex lethal, HuD, Hrp1, and nucleolin the two RRMs sandwich the RNA (Figure 1.6c and d).[18] In some cases, the resulting protein conformation induces several intra-RNA interactions that contribute to the overall enthalpy of the complex, while in other cases no intra-RNA interactions are present.[18] These different conformations may contribute to the specificity of the protein.[24] However, it is interesting that three RRM proteins, Sex lethal, HuD and Hrp1 have a very similar orientation, while the bound RNAs are different both in sequence and length.[24]

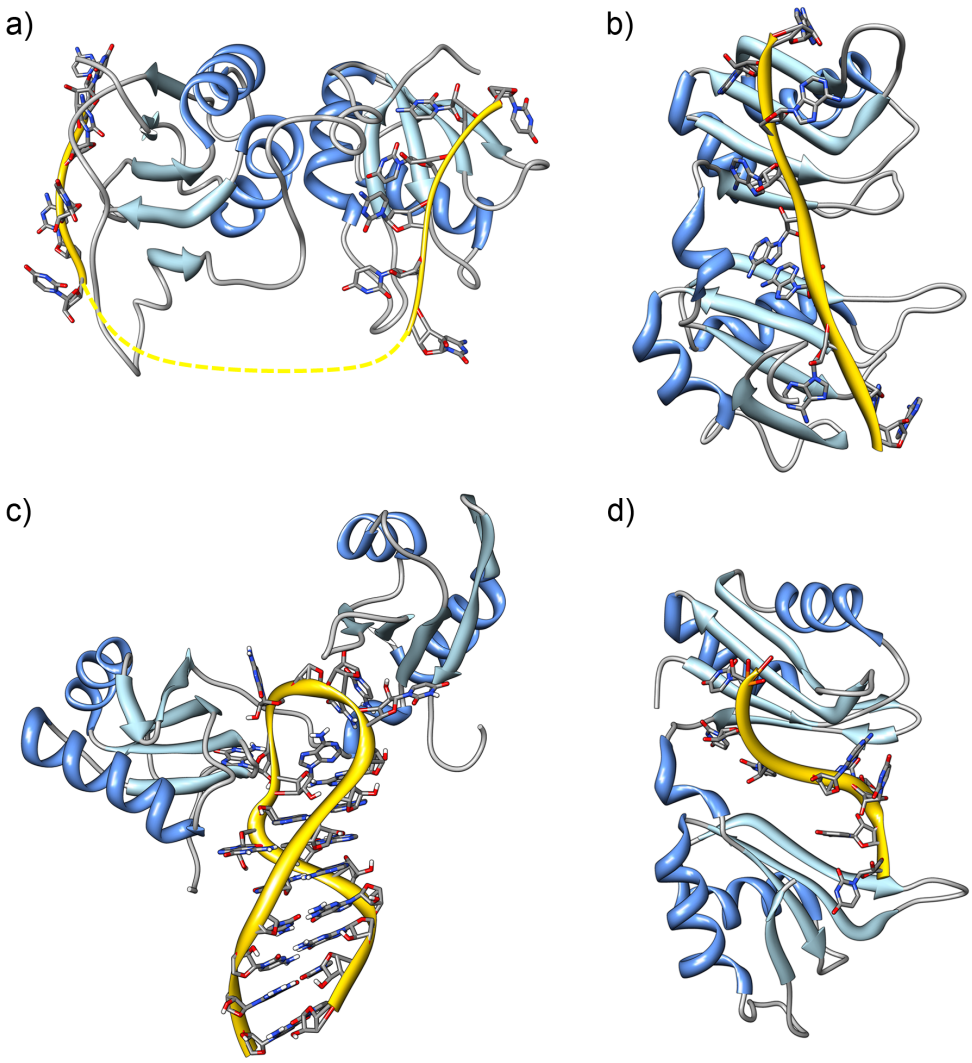



**Figure 1.6** Structure of mutiple RRM containing proteins in complex with RNA. a) RRM3 and RRM4 of PTB in complex with RNA. b) RRM1 and RRM2 of PABP in complex with RNA. c) RRM1 and RRM2 of nucleolin in complex with RNA. d) RRM1 and RRM2 of HuD in complex with RNA. These structures were generated with the program Chimera.[16]

*1.2.5 RRM-RRM interactions*

Both intramolecular RRM-RRM and intermolecular RRM-RRM interactions have been observed. RRMs within a protein may interact to form RNA binding surfaces. This is shown in PTB, as described in section 1.2.4, and Sex lethal. In the case of Sex lethal, the interaction

between the two RRMs clamp the RNA in a cleft formed between the two RRMs. The RNA binding is thought to induce the RRM-RRM interactions. The structure and the specific interactions that form in the complex will be discussed in section 1.4.2. RRM-RRM interactions have also been found in a homodimer-ssDNA complex. Two FIR proteins that are bound to the ssDNA are dimerized using RRM1 and RRM2 of each protein. While RRM1 recognizes the DNA, RRM2 participates in protein-protein contacts.[40] It is interesting that RRMs can interact with RRMs via both the RNA binding surface and the non-RNA binding surface. The RRM of SRSF1, an alternative splicing regulator, interacts with the RRM of U1-70K using the opposite side of the RNA binding surface.[19] In contrast, two proteins–Prp24 and Nup35–found in *S. cerevisiae* form RRM-RRM interaction with the RNA binding surface. In Prp24, a splicing factor that binds to U6 snRNA, RRM1 and RRM2 interact with each other. It is not clear how this protein binds to its target RNA since both of the canonical RNA binding surface is occupied by RRM-RRM interaction.[24] Nup35 protein, which interacts with the nulcear envelope, forms a homodimer using the β-sheet surface that contains atypical RNP sequences.

### *1.3 U1A protein*

The U1 snRNP specific polypeptide A (U1A) is a component of U1 snRNP and binds to the stem loop 2 (SL2) of U1 snRNA.[41] It also binds to the polyadenylation inhibition element (PIE) of its own pre-mRNA to regulate its expression.[42] U1A contains two RRMs, yet only the N-terminal RRM interacts specifically with SL2 RNA and PIE RNA.[42-44] The crystal structure of the N-terminal RRM of U1A in complex with the SL2 RNA was the first solved structure of an RRM in complex with RNA.[26] Thus U1A-SL2 complex has been extensively used as a model

system for studying the biochemical and thermodynamic properties of RRM-RNA interactions.[45-54]

### *1.3.1 Function of U1A protein*

#### *1.3.1.1 U1A in splicing*

U1A was first identified as a component of U1 snRNP, which interacts with U2, U4, U5 and U6 snRNP to form the major (U2-type) spliceosome (Figure 1.7a). U1 snRNP recognizes the 5' splice site and interacts with U2 snRNP. U1 snRNP consists of seven Sm proteins and U1snRNP specific U1-70k, U1A and U1C. U1C is crucial for the splicing activity of the spliceosome.[55] U1C aids the U1 snRNA to base pair with the 5' splice site.[56] However the function of U1-70k and U1A in splicing remain unclear.[57] In fact, *in vitro* splicing was not affected by the depletion of U1A[55] or the deletion of the stem loop 2 of U1 snRNA.[56] Recently, Shao *et al*. have reported that U1A is involved in the communication of U1 snRNP and U2 snRNP in fission yeast *Schizosaccharomyces pombe*. U1A binds to Rsd1, which interacts with Prp5. Prp5 then contacts SF3b of U2 snRNP (Figure 1.7c).[58] More biochemical studies are required to understand the exact role of U1A during spliceosome assembly and why *in vitro* splicing is not affected by the depletion of U1A or the deletion of the stem loop 2 of U1 snRNA.

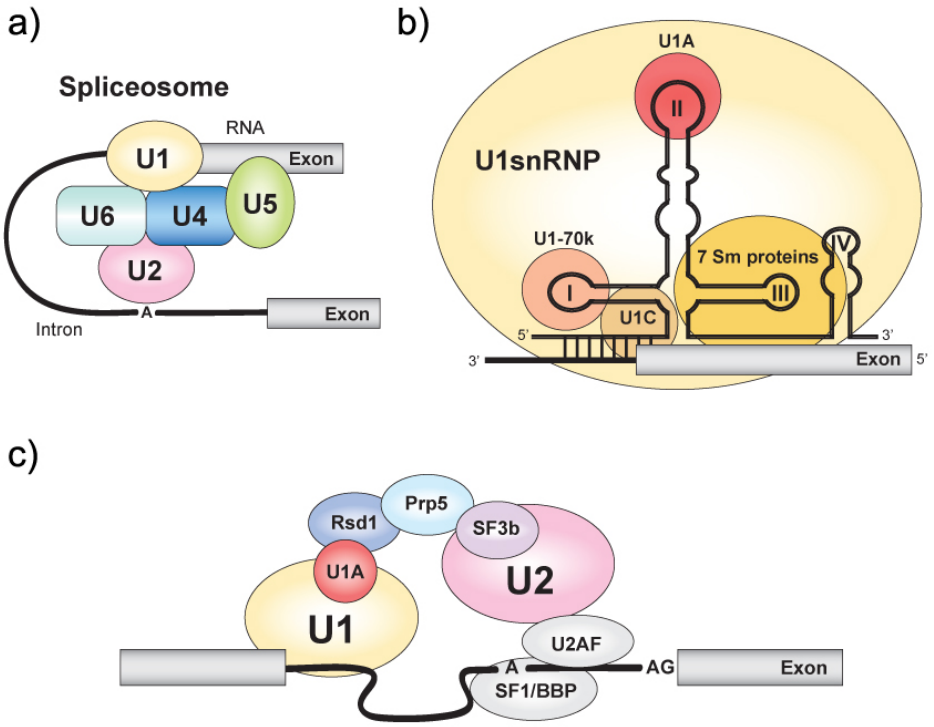



**Figure 1.7** U1A in the spliceosome. a) Schematic representation of spliceosome. Pre-mRNA splicing takes place within the spliceosome, which is composed of four snRNPs (U1, U2, U4/U6, and U5). b) Schematic representation of U1snRNP. U1snRNP binds to the 5' splice site of the target pre-mRNA. U1A binds to stem loop 2 of U1snRNA.[59] c) Protein interaction network of U1snRNP-U2snRNP communication during pre-spliceosome assembly. U1A protein binds to SR-like protein, Rsd1. Prp5 binds to Rsd1 and SF3b to connect U1snRNP and U2snRNP.[58]

*1.3.1.2 Autoregulation of its own pre-mRNA*

In higher eukaryotes, formation of the 3' end of most mRNAs is accomplished by endonucleolytic cleavage of the pre-mRNA followed by polyadenylation.[60] U1A regulates its own production by inhibiting the polyadenylation of U1A pre-mRNA.[42,61,62] Two U1A proteins cooperatively bind to the polyadenylation inhibition element (PIE) in the 3'-untranslated region of its own pre-mRNA.[63] Binding of the two U1A proteins to the PIE RNA enables the U1A protein to form a surface that interacts with the enzyme responsible for the synthesis of the poly(A) tail, poly(A) polymerase (PAP), and inhibits its activity.[63-65] The U1A-PIE trimolecular

complex does not block the binding of CPSF nor the cleavage reaction, indicating the specificity of the inhibition mechanism for the polyadenylation step.[66] Without polyadenylation, U1A pre-mRNA cannot form a mature 3' end therefore is degraded.

*1.3.1.3 Regulation of polyadenylation of other pre-mRNAs*

U1A also controls the polyadenylation of other pre-mRNAs, including pre-mRNAs of IgM heavy-chain.[67-70] Immunoglobins can occur in two forms, a soluble form that is secreted from the cell and a membrane-bound form that is attached to the surface of a B cell. U1A promotes the expression of the membrane-bound form of IgM heavy chain by preventing the polyadenylation at the secretory poly(A) site (Figure 1.8a). U1A binds to sequences (A(U/G)GCN$_{1-3}$C) that resemble the consensus U1A binding site, upstream and downstream of the poly(A) site. U1As that are bound to the three U1A binding motifs upstream of poly(A) secretory site, inhibit the polyadenylation using the same mechanism as the autoregulation process.[67,68] Whereas, U1As that are bound to the two U1A binding motifs downstream of the poly(A) site, occlude the binding of cleavage stimulatory factor (CstF) to the downstream GU-rich motifs and inhibit the formation of the cleavage/polyadenylation complex.[69] The N-terminal RRM of U1A binds to these novel sites, but with a 10-fold lower affinity than SL2 of U1 snRNA. It is suggested that this lower binding affinity is particularly suited to a regulatory role of U1A at this poly(A) site, which is fine-tuned by competing weak interactions.[67] During B cell differentiation the expression of the secretory form of IgM increases. Investigation of the U1A expression level revealed that the lower U1A availability is responsible for changes in its inhibitory effect at the secretory IgM poly(A) site.[70] Another experiment has shown that the level of the U1 snRNP associated U1A decreases upon B-cell stimulation.[71] Therefore U1

snRNP may be the functional unit for the regulation of IgM polyadenylation. Expression of secretory poly(A) site of other immunoglobins may also be regulated by U1A as implied by sequence anaylsis.[67]

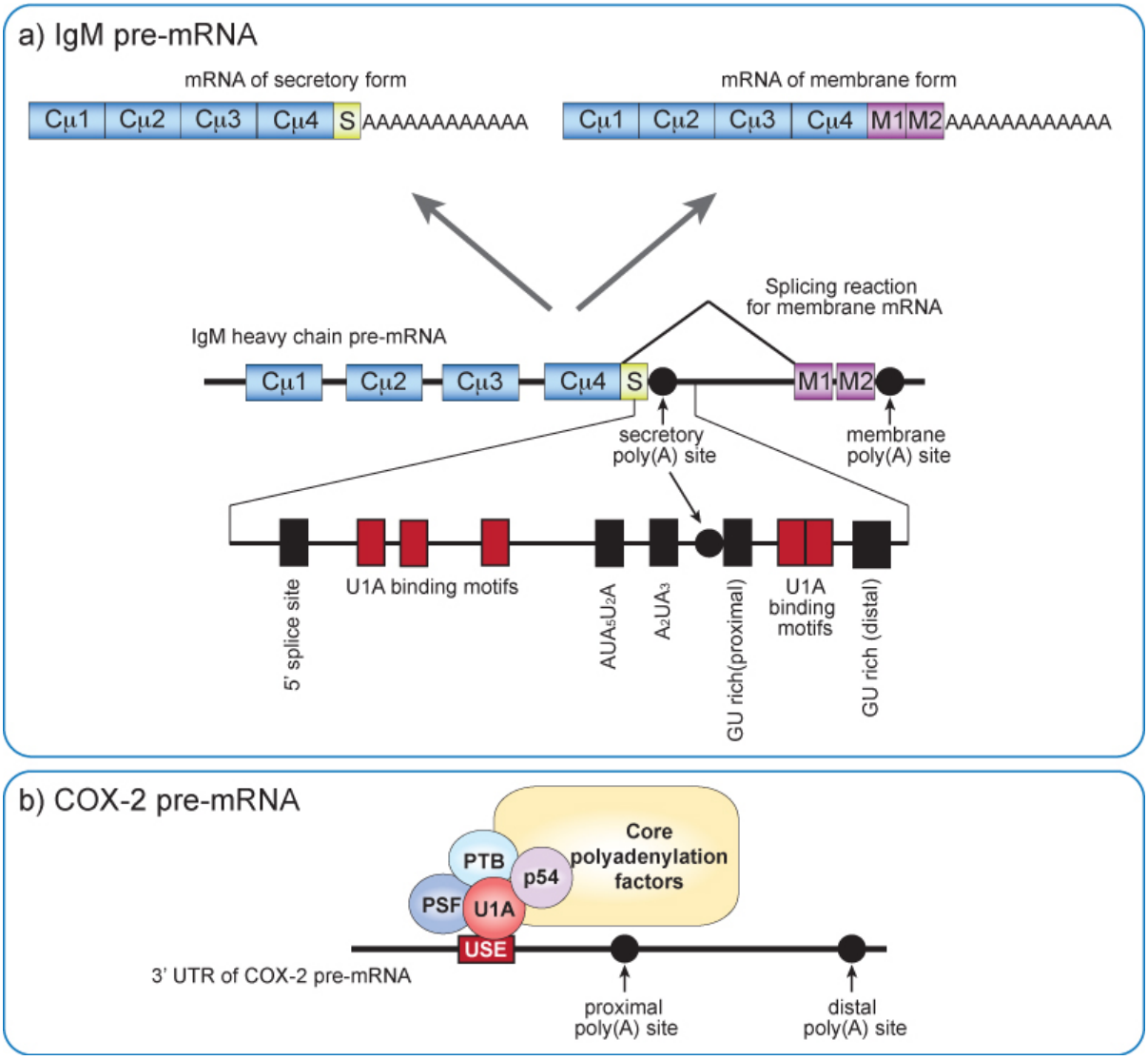



**Figure 1.8** Regulation of polyadenylation by U1A. a) Polyadenylation of IgM pre-mRNA. U1A binds to five U1A binding motifs and inhibits polyadenylation of secretory poly(A) site.[67] b) Polyadenylation of COX-2 pre-mRNA. A multi-protein complex including U1A directs the use of proximal poly(A) site.[72]

Regulation of polyadenylation by U1A may be accomplished by forming different multi-protein complexes. At least in two cases, U1A has been found to form multi-protein complexes independent of U1 snRNP. It is estimated that about 3% of U1A proteins in the cell are not associated with U1 snRNP.[73] Polyadenylation of Simian virus 40 (SV40)[73-75] and Cyclooxygenase-2 (COX-2)[72,76] pre-mRNAs is regulated by U1A and other proteins that interact with U1A. In the case of COX-2, two polyadenylation signals (PASs) are present in the 3' UTR; proximal and distal polyadenylation signals.[76] A protein complex that consists of U1A, $p54^{nrb}$, PSF and PTB binds to the *cis*-acting upstream sequence elements (USEs) upstream of proximal polyadenylation signal and promotes the usage of the proximal polyadenylation signal (Figure 1.8b).[72] The interactions of these proteins were not affected by the presence of RNase. Therefore the complex formation is not mediated by an RNA.[72,75] A detailed mechanism of how this complex induces the polyadenylation at the proximal polyadenylation site is yet to be determined.

*1.3.2 Structure of U1A protein and U1A-RNA complexes*

The human U1A is a 32 kDa protein comprised of two RRMs joined by a flexible linker rich in proline and lysine. Among the two RRMs only the N-terminal RRM (RRM1) interacts specifically with SL2 RNA of U1 snRNA and PIE RNA of its own pre-mRNA. No RNAs have been shown to bind to the C-terminal RRM (RRM2).[43,77] The structures of the free RRM1, free RRM2, RRM1-SL2 RNA complex, and RRM1-PIE RNA have been solved by X-ray crystallography and NMR.[26,28,63]

*1.3.2.1 U1A-SL2 RNA complex*

N-terminal RRM of U1A recognizes the single stranded element AUUGCAC in the stem loop 2 of U1 snRNA and contacts the GC base pair that closes the stem loop. U1A binds weakly to a RNA that has the same consensus sequence but does not form a stem loop. The three residues (UCC) that are in the loop but do not contact the RNA are flexible and can be randomized or replaced by a polyethylene glycol spacer.[78,79]

RRM1 of U1A follows the basic binding mode utilizing the conserved aromatic residues in β1 and β3. Residues that correspond to position 2 of RNP2 and position 5 of RNP1 are Tyr13 and Phe56, which form stacking interactions with C10 and A11, respectively. Position 3 of RNP1 is a glutamine that forms a stacking interaction with G9 of the RNA. Glu54 also forms hydrogen bonds with the hydroxyl group of Tyr13 positioning it to form the stacking interaction with C10.

Loop 1 and loop 3 actively participate in recognition of the RNA. Asn16, Leu17, Glu19 in loop 1 forms direct or water mediated hydrogen bonds with the RNA. Loop 3 (Ser46-Arg52) is flexible in the free protein but becomes fixed in the complex. The loop protudes into the RNA loop presenting the bases of the RNA to the β-strands of the protein. Arg52 plays a crucial role in binding. Mutation of Arg52 to glutamine abolishes RNA binding. Arg52 not only contacts two nucleotides (A6 and G16; Figure 1.5a) but also forms a hydrogen bond with two residues in loop 3 (Arg47 and Ser48) defining the conformation of loop 3.

Helix C is in a closed form contacting the β-surface of the RRM1 in the free protein and is moved away from the β-surface exposing the β-surface for RNA binding. When the protein is bound to the RNA, residues in helix C directly contact the RNA by forming direct or water mediated hydrogen bonds (Figure 1.9c).

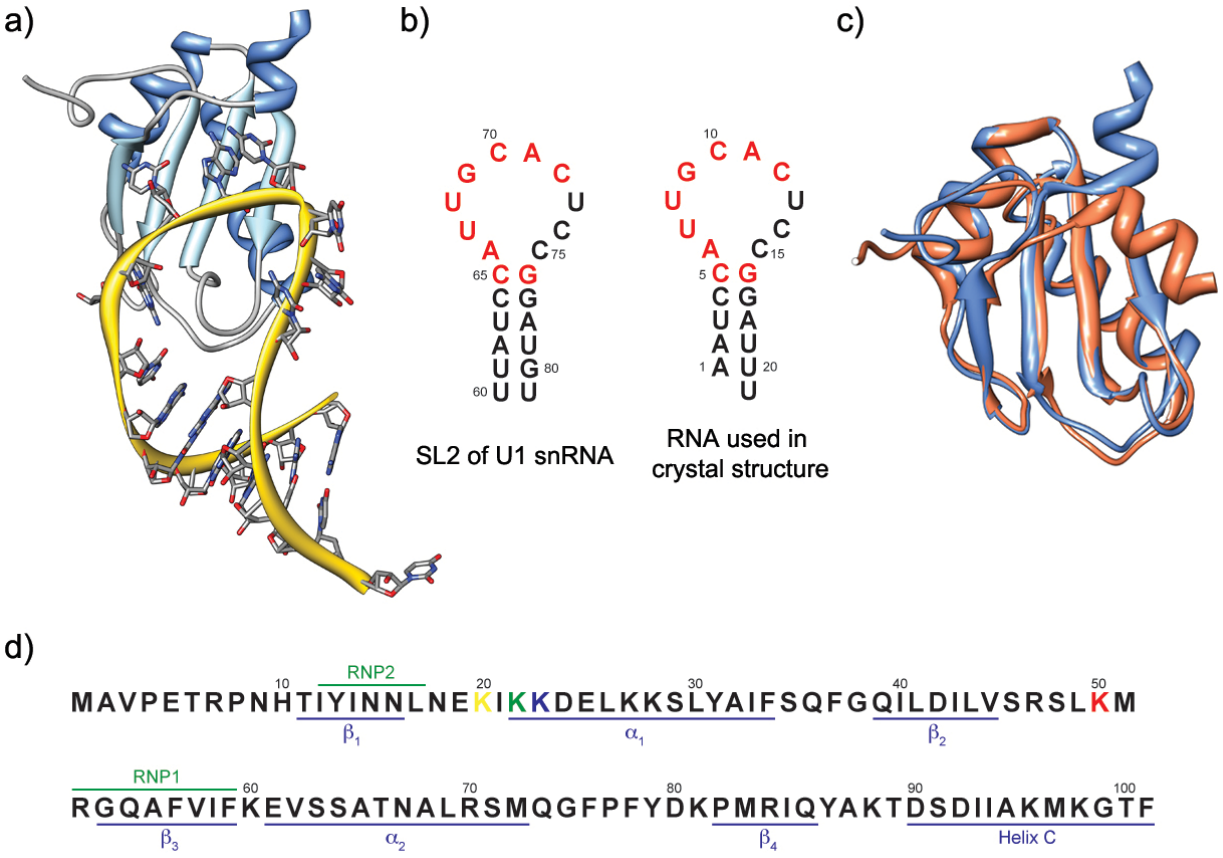



**Figure 1.9** a) Structure of U1A-SL2 RNA. b) Sequences of SL2 RNA. c) Structure alignment of free U1A (blue) and RNA bound U1A (orange). d) Primary sequence of N-terminal RRM of U1A protein. These structures were generated with the program Chimera.[16]

*1.3.2.2 U1A-PIE RNA complex*

The 3' untranslated region of U1A pre-mRNA folds into a conserved secondary structure with two internal loops; one loop contains the consensus sequence (AUUGCAC) and the other its variant (AUUG*U*AC) (Figure 1.10a).[61] Two U1A proteins bind cooperatively to two internal loops and form a trimolecular complex. This trimolecular complex was proposed using a modeling program[80] and verified by an NMR structure with a similar RNA sequence (Figure 1.10b).[63] Although the structure of the PIE RNA is quite different from the SL2 RNA, the U1A-PIE complex is formed using the same sequence-specific interactions between the RNA and the

protein with slight variation in the helix C-RNA interface. The cooperativity of the two U1A molecules is directed by protein-protein interactions and not by protein-induced changes in the RNA.

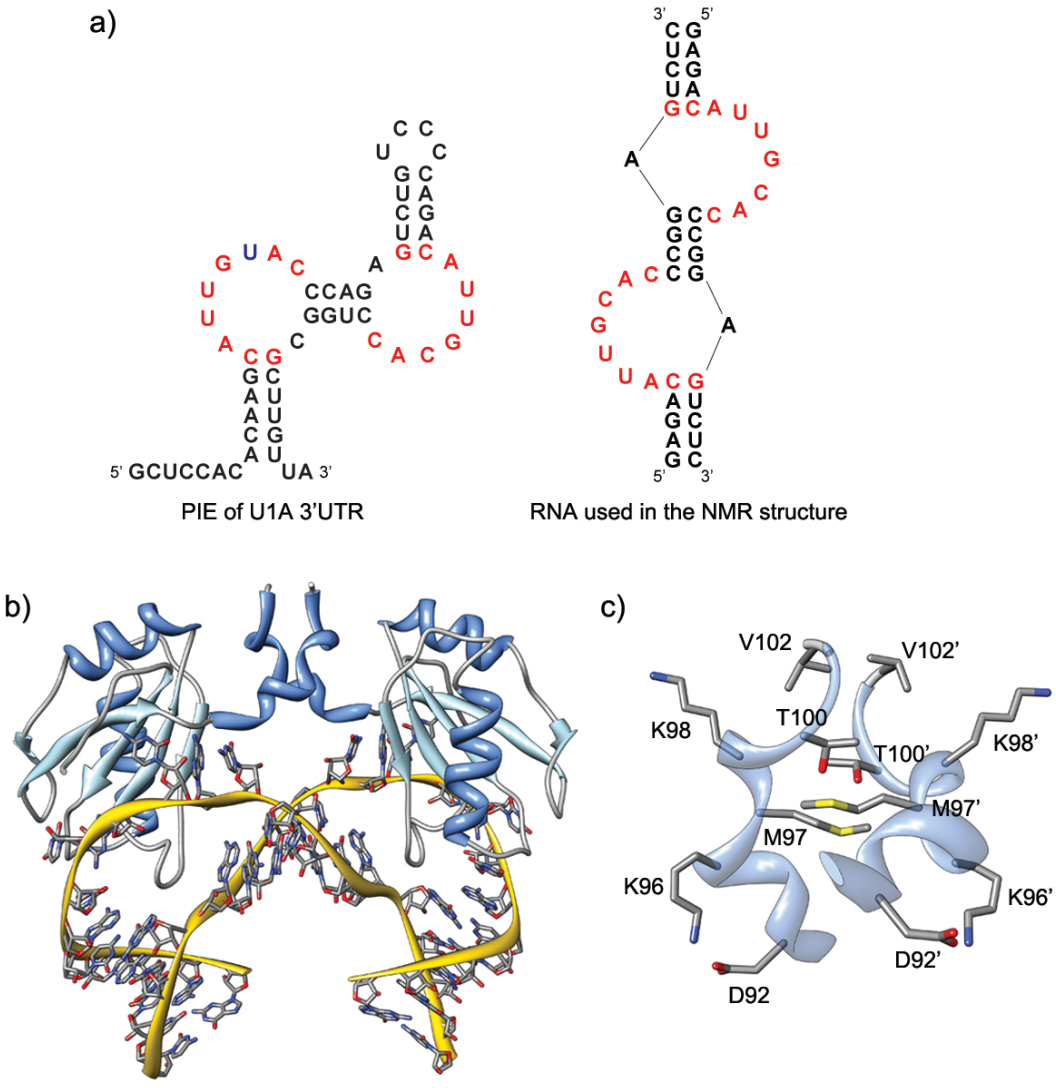



**Figure 1.10** a) Sequence of PIE RNA and RNA used in the NMR structure. Sequences that directly contact U1A are highlighted in red. b) Structure of the U1A-PIE complex. (pdb: 1DZ5) c) Close-up view of helix C-helix C interaction. These structures were generated with the program Chimera.[16]

The protein-protein interaction is mediated by helix C. Helix C is amphiphatic, with hydrophilic residues (Asp92, Lys96, Lys98) pointing to the RNA or exposed to the solvent and hydrophobic residues (Met97, Thr100, Val102) participating in protein-protein interactions. Met97 is at the core of the helix C interaction, packed against Ile93 and contacting Gly99, Thr100 and the other Met97 (Figure 1.10c). Due to these interactions helix C becomes more rigid.[30] The sequence following helix C contains the PAP binding site, which spans residues 103-115.[64] This sequence contains conserved basic and acidic amino acids and mutations of some of the basic amino acids reduce inhibition of PAP.[65] Based on the NMR structure, the PAP binding region may be formed by an extension of helix C.

### *1.4 Sex lethal protein*

The Sex lethal protein is a master switch for determining the sex of *Drosophila Melanogaster*. Produced only in female flies, Sex lethal directs all aspects of female development and prevents the activation of the male-specific dosage compensation system. In males, *Sxl* is OFF, resulting in male development and activation of X chromosome dosage compensation. Sex lethal regulates alternative splicing and translation of pre-mRNAs that are involved in sex development and dosage compensation. These post-transcriptional regulation processes are accomplished by the protein's ability to specifically bind to its target RNAs. Sex lethal contains two RRMs and, unlike U1A, it requires both RRMs to bind RNA.

*1.4.1 Sex lethal and sex determination of Drosophila Melanogaster*

*1.4.1.1 Expression of Sex lethal: Activation and maintenance*

*Sxl* contains two promoters: establishment (*Pe*) and maintenance (*Pm*).[81,82] The establishment promoter is temporarily available during early embryogenesis and is permanently shut down at the cellular blastoderm stage.[82,83] The use of establishment promoter depends on the number of X chromosomes, which leads to an all-or-none response of *Sxl*.[83,84] Four X-encoded proteins, collectively called X-linked signal elements (XSE), serve as the primary determinants of X chromosome dose.[82] Three of the XSE proteins–SisA, Scute and Runt–are transcription factors that stimulate the transcription of *SxlPe*.[85] Another XSE protein, Unpaired activates the maternally supplied *Stat92E* transcription factor, which in turn reinforces the activation of *SxlPe*.[86-88] In female flies (XX), the concentrations of these four XSE proteins reach a certain threshold level to activate the transcription of *SxlPe* (Figure 1.11a).[83] Whereas in males (XY), XSE proteins never exceed threshold levels and *Pe* remains inactive (Figure 1.11b). Once the X chromosome dose has been assessed, the establishment promoter is inactivated and *Sxl* is primarily transcribed by the maintenance promoter.[82] The transcript from the maintenance promoter contains a male-specific exon, which includes multiple premature stop codons. In females, Sex lethal forces this exon to be skipped, therefore producing the full-length functional protein (Figure 1.11c). In males, absence of Sex lethal leads to the inclusion of the male-specific exon, resulting in truncated non-functional proteins (Figure 1.11d). Thus, the autoregulatory splicing of Sex lethal converts the transient X chromosome dose signal into long-term cellular memory.

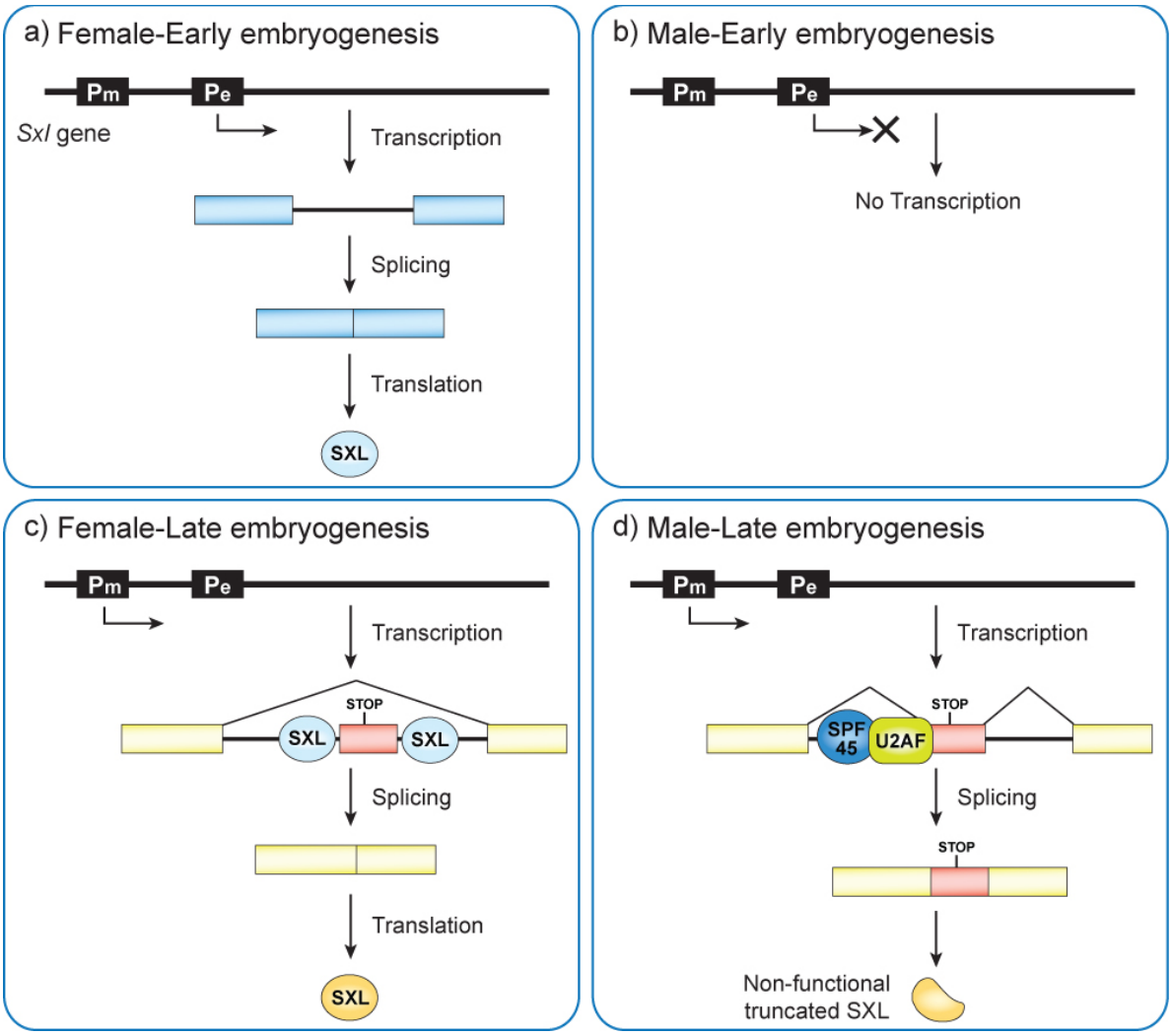



**Figure 1.11** Expression of Sex lethal in *Drosophila Melanogaster*. a) Early expression of Sex lethal in female flies. b) In male flies, the establishment promoter is not used and results in no transcription. c) Once the early Sex lethal is expressed, it regulates alternative splicing of its own pre-mRNA resulting in the translation of Sex lethal. d) In males, the absence of Sex lethal results in a transcript that includes multiple stop codons in the middle of the transcript, which is transcribed into a non-functional truncated Sex lethal.

Two Sex lethal binding sites have been identified in its own pre-mRNA. One is >200 nucleotides downstream and the other is >200 nucleotides upstream of the male-specific exon.[89,90] Sex lethal proteins bound to these sites are proposed to interact with general splicing

factors (U2AF and SPF45) and components of U1 snRNP (SNF and U1-70k) to inhibit their function.[91] Recent studies have suggested that other proteins may also promote exon skipping including Transformer, which is a downstream target of Sex lethal.[92]

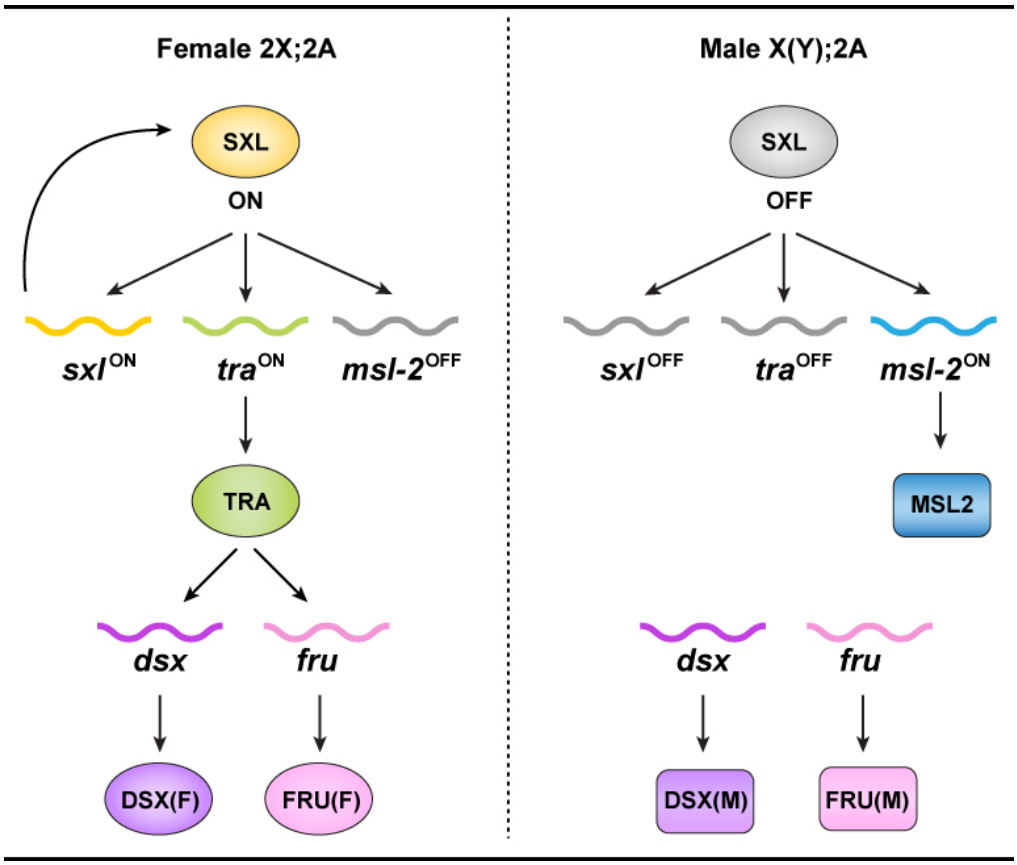



**Figure 1.12** Sex determination pathway of *D. Melanogaster*. The X chromosome dose signal targets the *Sxl* gene, establishing the autoregulation of *Sxl* in embryos. *Sxl*, *tra*, *msl-2*, *dsx*, and *fru* genes are all differentially spliced in males and females. Sex lethal protein turns on *Sxl* itself and *tra* via alternative splicing of the corresponding pre-mRNAs. Transformer then activates female-specific splicing of *dsx* and *fru*. The absence of Transformer protein in male results in male-specific protein variants of *dsx* and *fru* that regulate distinct sets of target genes. The production of Msl-2 protein is prevented by Sex lethal in female flies.

*1.4.1.2 Splicing and translational regulation of target genes*

Sex lethal is at the top of the sex determination pathway, orchestrating sex-specific development and behavior by modulating the expression of a set of downstream genes (Figure 1.12).[82,93] Sex lethal has been found to regulate the expression of at least two target genes–*transformer* (*tra*) and *male-specific lethal-2* (*msl-2*)–aside from *Sxl* itself.[82]

*Tra* is the primary effector through which *Sxl* controls sexual differentiation.[82,94] Transformer is an RNA binding protein that activates female-specific splicing of *doublesex* (dsx) and *fruitless* (*fru*) genes.[95] Sex lethal protein controls alternative splicing of *tra* pre-mRNA by binding to the polypyrimidine tract associated with the non-sex specific splice site and diverting U2AF, a general splicing factor to the downstream polypyrimidine tract.[96] Thus U2AF recruits the splicing machinery to the female-specific splice site (Figure 1.13). Use of the proximal 3' splice site introduces a stop codon in the open reading frame leading to the translation of a non-functional protein while the use of distal 3' splice site results in a functional Transformer protein. In females, approximately half of the *tra* pre-mRNA is spliced at the downstream splice site to produce Transformer protein.[97,98]

In males, X chromosome dosage compensation complex is formed, which increases the transcriptional output of the X chromosome. Msl-2 is a component of this dosage compensation complex required for its function. In females, the expression of Msl-2 is repressed by Sex lethal thereby inhibiting the formation of the dosage compensation complex. Sex lethal not only regulates the splicing of *msl-2* in the nucleus but also inhibits the translation of the transcript in the cytoplasm. *Msl-2* RNA contains multiple Sex lethal binding sites at both the 5' and 3' untranslated regions (UTR). At the 5' UTR, Sex lethal binds to both the 5' splice site and the 3' splice site of the first intron and prevents the splicing of this intron. At the 5' splice site, Sex

lethal displaces Rox8, which enhances the use of the weak 5' splice site. At the 3' splice site, Sex lethal blocks the binding of U2AF as in *tra* RNA (Figure 1.13). Interestingly, the retention of this intron does not have an effect on the final protein product since it is not in the open reading frame.[99,100] Instead, Sex lethal proteins bound to the intron are used to suppress the translation of *msl-2* by inhibiting the scanning of the 43S ribosome subunit. In addition to this mechanism, Sex lethal proteins bound at the 3' UTR block the binding of 43S complex to the 5' end of the mRNA. This is achieved by recruiting a co-repressor, Upstream of N-ras (UNR), which in turn binds to PABP (Figure 1.13).[101-104] However the exact mechanism of 43S complex inhibition needs further investigation.

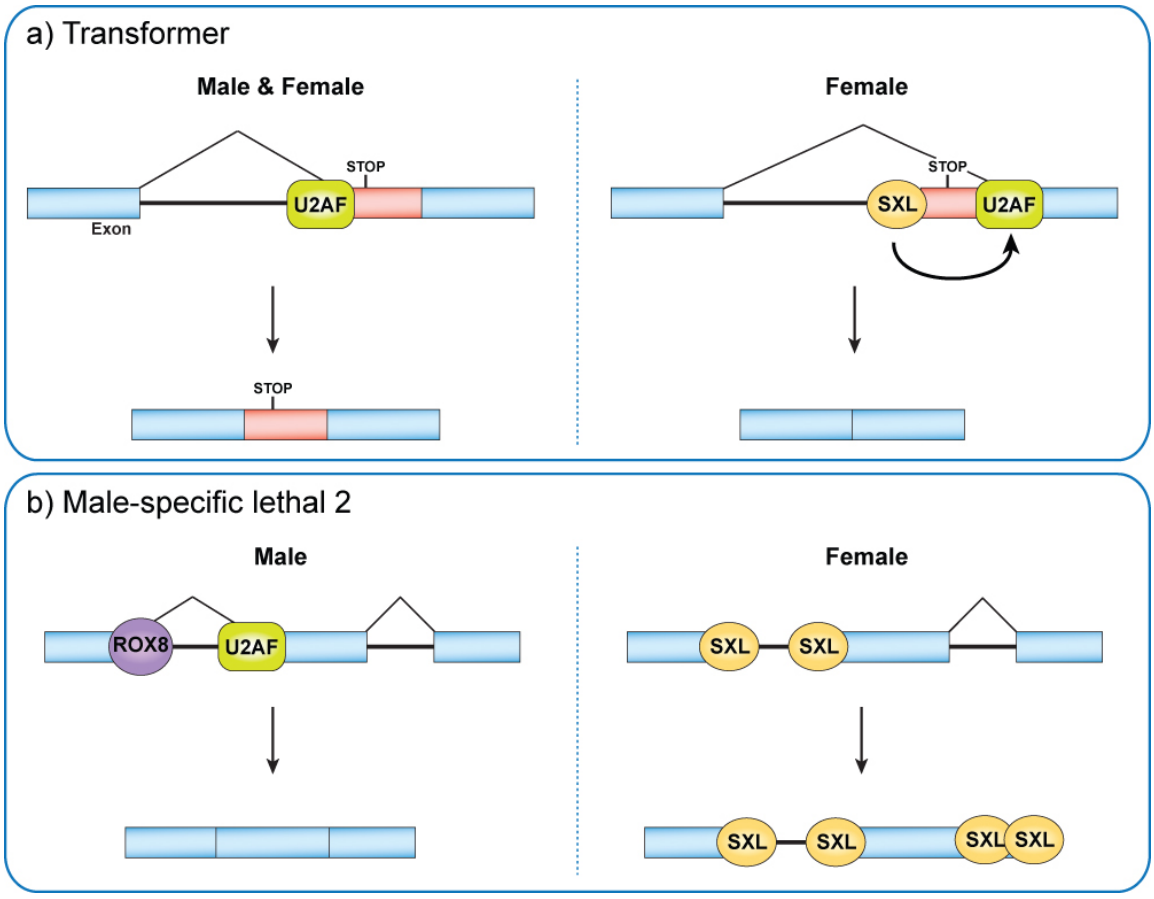



**Figure 1.13** Splicing regulation by Sex lethal. a) *tra* RNA. b) *msl-2* RNA.[94]

The exact number of target genes of Sex lethal is not known. Recent bioinformatic approaches[105] and biochemical studies[106,107] have identified a few more plausible targets. For example, Sex lethal down-regulates Notch protein, which controls the bristle number on the adult cuticle.[107] *Notch* mRNA contains several Sex lethal binding sites in its 5' and 3' UTRs and Sex lethal is capable of binding *Notch* mRNA. The number and distribution of the Sex lethal binding sites are similar to that of *msl-2*. However whether the mechanism of translational repression is the same as that of *msl-2* is not known.[107]

*1.4.2 Structure of Sex lethal-tra RNA complex*

Sex lethal is a 354 amino-acid protein that binds to polypyrimidine tracts in mRNAs using two RNA recognition motifs. Unlike U1A, Sex lethal requires both RRMs for it to bind to the target sequence. Like many of the RRM proteins, Sex lethal binds to unstructured single stranded RNAs. Biochemical analysis using SELEX revealed that Sex lethal binds to a consensus sequence of UUUUGUU(G/U)U(G/U)UUU(G/U)UU.[108] Based on this consensus sequence, genome analysis identified more than 10,000 potential Sex lethal binding sites in Drosophila genome.[105]

Crystal and NMR structures of Sex lethal free and bound to the polypyrimidine tract (GUUGUUUUUUUU) of *tra* RNA have been solved.[36] The overall structure of the complex is shown in Figure 1.14a. The two RRMs have their β-sheet surfaces facing each other forming a V-shaped cleft. Sex lethal recognizes nine nucleotides (U3-U11) continuously. The cleft is strongly electropositive and binds to six nucleotides in the 3' end (U6-U11). RRM2 contacts three nucleotides (U3-G4-U5) close to the 5' end. The bound RNA has a turn in the middle, which is characterized by an intramolecular stacking interaction between U7 and U8. The main

difference between the structure of the free protein and the complex comes from the interdomain linker. According to NMR studies and crystal structure of the free protein, the linker between the two RRMs is unstructured and flexible without the RNA.[109,110] However upon RNA recognition it forms a rigid $3_{10}$ helix and forms hydrogen bonds with RRM2. In addition to the conformational change of the linker, two electrostatic contacts form between the two RRMs that are not observed in the free protein. The side chain of Lys197 contacts the main chain carbonyl group of Val238 and the hydroxyl group of Tyr131 contacts the side chain of Glu239 (Figure 1.14b).

The conserved RNP sequences of RRM1 and RRM2 participate in stacking interaction with the RNA bases (Figure 1.14c). Residues of RNP1 position 5 in both RRMs are phenylalanines that form stacking interactions with U11 and U5 respectively. In RRM2, the third residue of RNP1 is Val254 and the second residue in RNP2 is Tyr214, which each form stacking interactions with U3 and G4. The corresponding residues in RRM1 do not form stacking interactions with any bases. Instead, position 4 and 5 of RNP2 in RRM1–Asn130 and Tyr131–forms stacking interaction with U9 and U6. Overall, six stacking interactions, which are mediated by residues in the two RNP sequences, play a significant role in the binding affinity of the protein-RNA complex. Similar to most RRM proteins, these residues do not contribute to the specificity of the protein.

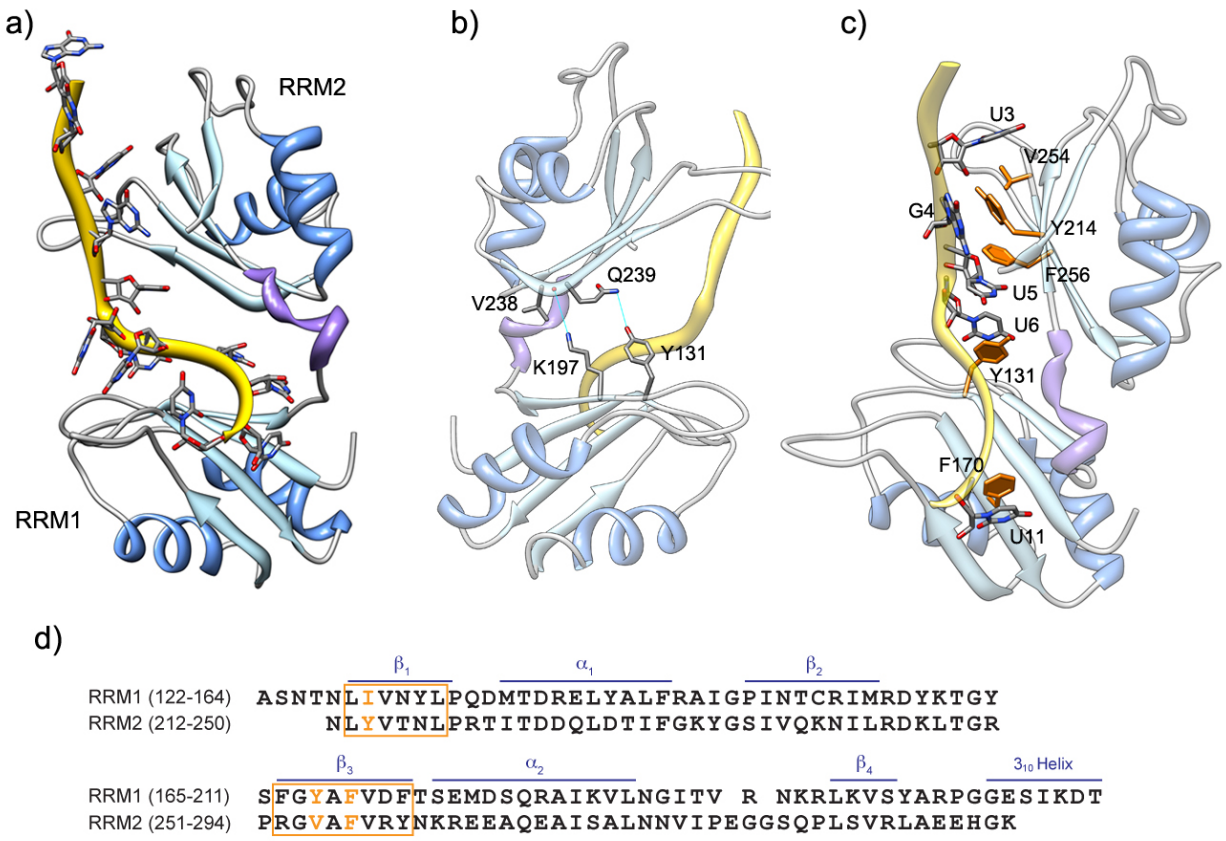



**Figure 1.14** Structure of Sex lethal-*tra* RNA complex. a) Overall structure of the complex. α helices are in blue, β strands are in light blue, $3_{10}$ helix in the interdomain linker is in purple and the RNA backbone is in yellow. b) RRM-RRM interaction. c) Stacking interactions formed between Sex lethal and RNA via residues in the RNP sequences. These structures were generated with the program Chimera.[16]

Residues in the loop regions as well as β strands are responsible for the specific recognition of the *tra* RNA. They form hydrogen bonds or have electrostatic interactions with the RNA bases (Table 1.2). All nine nucleotides (UGUUUUUUU) are specifically recognized by the Sex lethal. These specific interactions explain the mechanism by which uridine-rich sequences with one or more cytidine residues bind weakly to the Sex lethal.[33,111] However,

substitution of uracil with guanine is tolerated in some positions, presumably because uracil and guanine have similar functional displayed along the edge of the base.[33]

**Table 1.2** Salt bridges formed between Sex lethal and *tra* RNA bases.

| | RRM1 | RNA | RRM2 | RNA |
|---|---|---|---|---|
| β1 | Asn 126 | U11 | Asn 212 | U5 |
| | | | Asn 217 | U3 |
| Loop 1 | Gln 134 | U7 & U8 | | |
| β2 | Arg 155 | U11 | Gln 239 | U5 |
| Loop 3 | Ser 165 | U8 | Arg 252 | U3 |
| | Tyr 166 | U8 | | |
| Loop 4 | Arg 195 | U6 | | |
| β4 | Lys 197 | U9 | Ala 289 | G4 |
| C-terminal | Arg 202 | U9 & U10 | | |
| | Gly 204 | U10 | | |

Sex lethal does not bind to DNA or an RNA containing deoxyuridines.[33,34] Consistent with these binding studies, six of the nine 2'-OH groups are in contact with Sex lethal residues. Moreover, *tra* RNA contains intramolecular hydrogen bonds between the 2'-OH and phosphate groups in the complex. These contacts form the kink in the middle of the RNA that allows the protein to interact with RNA bases.

### *1.5 Summary*

RNA binding proteins play an essential role in all steps of gene expression and gene regulation. RNA binding proteins are built from a small number of RNA-binding domains to create a diverse set of RNA binding proteins many of which remain to be discovered and characterized. The RNA recognition motif is an abundant and diverse RNA binding domain found in all organisms. Biochemical studies and structural analysis of RRM-containing proteins have shown that this small domain is extremely diverse in terms of both function and structure.

The U1A protein is not only found as a component of the U1 snRNP but also in a distinct complex that regulates polyadenylation of several pre-mRNAs. Although U1A contains two RRMs, only the N-terminal RRM interacts specifically to bind RNA. Extensive structural and biochemical studies have been performed on the U1A-SL2 RNA complex making it an ideal model system to study RRM-RNA interactions. The Sex lethal protein is an alternative splicing regulator that determines the sex of *Drosophila Melanogaster*. Sex lethal binds to unstructured single-stranded RNA using both RRMs. In the following two chapters, Sex lethal-*tra* RNA and U1A-SL2 RNA complexes are used as a model system for the identification of small molecule inhibitors of RRM-RNA interactions and to understand the kinetics of RRM-RNA complex dissociation.

### *1.6 References*


1 Sharp, P. A. The centrality of RNA, *Cell* **2009**, *136*, 577.

2 Bayne, E. H.; Allshire, R. C. RNA-directed transcriptional gene silencing in mammals, *Trends Genet* **2005**, *21*, 370.

3 Fedor, M. J.; Williamson, J. R. The catalytic diversity of RNAs, *Nat Rev Mol Cell Biol* **2005**, *6*, 399.

4 Matera, A. G.; Terns, R. M.; Terns, M. P. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs, *Nat Rev Mol Cell Biol* **2007**, *8*, 209.

5 Dreyfuss, G.; Kim, V. N.; Kataoka, N. Messenger-RNA-binding proteins and the messages they carry, *Nat Rev Mol Cell Biol* **2002**, *3*, 195.

6 Glisovic, T.; Bachorik, J. L.; Yong, J.; Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation, *FEBS Lett* **2008**, *582*, 1977.

7 Keene, J. D. RNA regulons: coordination of post-transcriptional events, *Nat Rev Genet* **2007**, *8*, 533.

8 Konig, J.; Zarnack, K.; Luscombe, N. M.; Ule, J. Protein-RNA interactions: new genomic technologies and perspectives, *Nat Rev Genet* **2011**, *13*, 77.

9 Kishore, S.; Luber, S.; Zavolan, M. Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression, *Brief Funct Genomics* **2010**, *9*, 391.

10 Lasko, P. Gene regulation at the RNA layer: RNA binding proteins in intercellular signaling networks, *Sci STKE* **2003**, RE6.

11 Chaudhury, A.; Chander, P.; Howe, P. H. Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles, *RNA* **2010**, *16*, 1449.

12 Lestrade, L.; Weber, M. J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs, *Nucleic Acids Res* **2006**, *34*, D158.

13 Decatur, W. A.; Fournier, M. J. RNA-guided nucleotide modification of ribosomal and other RNAs, *J Biol Chem* **2003**, *278*, 695.

14 Kishore, S.; Stamm, S. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C, *Science* **2006**, *311*, 230.

15 van Kouwenhove, M.; Kedde, M.; Agami, R. MicroRNA regulation by RNA-binding proteins and its implications for cancer, *Nat Rev Cancer* **2011**, *11*, 644.

16 Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem* **2004**, *25*, 1605.

17 Lunde, B. M.; Moore, C.; Varani, G. RNA-binding proteins: modular design for efficient function, *Nat Rev Mol Cell Bio* **2007**, *8*, 479.

18 Maris, C.; Dominguez, C.; Allain, F. H. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression, *FEBS J* **2005**, *272*, 2118.

19 Cho, S.; Hoang, A.; Sinha, R.; Zhong, X. Y.; Fu, X. D.; Krainer, A. R.; Ghosh, G. Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly, *Proc Natl Acad Sci U S A* **2011**, *108*, 8233.

20 Naeeni, A. R.; Conte, M. R.; Bayfield, M. A. RNA chaperone activity of human la protein is mediated by variant RNA recognition motif, *J Biol Chem* **2012**, *287*, 5472.

21 LaBranche, H.; Dupuis, S.; Ben-David, Y.; Bani, M. R.; Wellinger, R. J.; Chabot, B. Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase, *Nat Genet* **1998**, *19*, 199.

22 Zhao, Y.; Baranger, A. M. Design of an adenosine analogue that selectively improves the affinity of a mutant U1A protein for RNA, *J Am Chem Soc* **2003**, *125*, 2480.

23 Auweter, S. D.; Fasan, R.; Reymond, L.; Underwood, J. G.; Black, D. L.; Pitsch, S.; Allain, F. H. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1, *EMBO J* **2006**, *25*, 163.

24 Clery, A.; Blatter, M.; Allain, F. H. T. RNA recognition motifs: boring? Not quite, *Curr Opin Struc Biol* **2008**, *18*, 290.

25 Ding, J.; Hayashi, M. K.; Zhang, Y.; Manche, L.; Krainer, A. R.; Xu, R. M. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA, *Genes Dev* **1999**, *13*, 1102.

26 Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C. H.; Nagai, K. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature* **1994**, *372*, 432.

27 Dominguez, C.; Allain, F. H. NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition, *Nucleic Acids Res.* **2006**, *34*, 3634.

28 Nagai, K.; Oubridge, C.; Jessen, T. H.; Li, J.; Evans, P. R. Crystal-Structure of the Rna-Binding Domain of the U1 Small Nuclear Ribonucleoprotein-A, *Nature* **1990**, *348*, 515.

29 Zeng, Q. Y.; Hall, K. B. Contribution of the C-terminal tail of U1A RBD1 to RNA recognition and protein stability, *RNA* **1997**, *3*, 303.

30 Mittermaier, A.; Varani, L.; Muhandiram, D. R.; Kay, L. E.; Varani, G. Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein, *J Mol Biol* **1999**, *294*, 967.

31 Perez Canadillas, J. M.; Varani, G. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein, *EMBO J* **2003**, *22*, 2821.

32 Deka, P.; Rajan, P. K.; Perez-Canadillas, J. M.; Varani, G. Protein and RNA dynamics play key roles in determining the specific recognition of GU-rich polyadenylation regulatory elements by human Cstf-64 protein, *J Mol Biol* **2005**, *347*, 719.

33 Singh, R.; Banerjee, H.; Green, M. R. Differential recognition of the polypyrimidine-tract by the general splicing factor U2AF65 and the splicing repressor Sex-lethal, *RNA* **2000**, *6*, 901.

34 Kanaar, R.; Lee, A. L.; Rudner, D. Z.; Wemmer, D. E.; Rio, D. C. Interaction of the Sex-lethal RNA binding domains with RNA, *EMBO J* **1995**, *14*, 4530.

35 Deo, R. C.; Bonanno, J. B.; Sonenberg, N.; Burley, S. K. Recognition of polyadenylate RNA by the poly(A)-binding protein, *Cell* **1999**, *98*, 835.

36 Handa, N.; Nureki, O.; Kurimoto, K.; Kim, I.; Sakamoto, H.; Shimura, Y.; Muto, Y.; Yokoyama, S. Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein, *Nature* **1999**, *398*, 579.

37 Oberstrass, F. C.; Auweter, S. D.; Erat, M.; Hargous, Y.; Henning, A.; Wenter, P.; Reymond, L.; Amir-Ahmady, B.; Pitsch, S.; Black, D. L.; Allain, F. H. Structure of PTB bound to RNA: specific binding and implications for splicing regulation, *Science* **2005**, *309*, 2054.

38 Lamichhane, R.; Daubner, G. M.; Thomas-Crusells, J.; Auweter, S. D.; Manatschal, C.; Austin, K. S.; Valniuk, O.; Allain, F. H.; Rueda, D. RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression, *Proc Natl Acad Sci U S A* **2010**, *107*, 4105.

39 Jenkins, J. L.; Shen, H.; Green, M. R.; Kielkopf, C. L. Solution conformation and thermodynamic characteristics of RNA binding by the splicing factor U2AF65, *J Biol Chem* **2008**, *283*, 33641.

40 Crichlow, G. V.; Zhou, H.; Hsiao, H. H.; Frederick, K. B.; Debrosse, M.; Yang, Y.; Folta-Stogniew, E. J.; Chung, H. J.; Fan, C.; De la Cruz, E. M.; Levens, D.; Lolis, E.; Braddock, D. Dimerization of FIR upon FUSE DNA binding suggests a mechanism of c-myc inhibition, *EMBO J* **2008**, *27*, 277.

41 Sillekens, P. T.; Habets, W. J.; Beijer, R. P.; van Venrooij, W. J. cDNA cloning of the human U1 snRNA-associated A protein: extensive homology between U1 and U2 snRNP-specific proteins, *EMBO J* **1987**, *6*, 3841.

42 Boelens, W. C.; Jansen, E. J.; van Venrooij, W. J.; Stripecke, R.; Mattaj, I. W.; Gunderson, S. I. The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA, *Cell* **1993**, *72*, 881.

43 Scherly, D.; Boelens, W.; van Venrooij, W. J.; Dathan, N. A.; Hamm, J.; Mattaj, I. W. Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA, *EMBO J* **1989**, *8*, 4163.

44 Lutz-Freyermuth, C.; Query, C. C.; Keene, J. D. Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA, *Proc Natl Acad Sci U S A* **1990**, *87*, 6393.

45 Varani, G.; Nagai, K. RNA recognition by RNP proteins during RNA processing, *Annu Rev Biophys Biomol Struct* **1998**, *27*, 407.

46 Allain, F. H.; Howe, P. W.; Neuhaus, D.; Varani, G. Structural basis of the RNA-binding specificity of human U1A protein, *EMBO J* **1997**, *16*, 5764.

47 Hall, K. B. Interaction of RNA Hairpins with the Human U1A N-Terminal RNA-Binding Domain, *Biochemistry* **1994**, *33*, 10076.

48 Katsamba, P. S.; Bayramyan, M.; Haworth, I. S.; Myszka, D. G.; Laird-Offringa, I. A. Complex role of the beta(2)-beta(3) loop in the interaction of U1A with U1 hairpin II RNA, *J Biol Chem* **2002**, *277*, 33267.

49 Katsamba, P. S.; Myszka, D. G.; Laird-Offringa, I. A. Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA, *J Biol Chem* **2001**, *276*, 21476.

50 Mittermaier, A.; Varani, L.; Muhandiram, D. R.; Kay, L. E.; Varani, G. Changes in sidechain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein (vol 294, pg 967, 1999), *J Mol Biol* **2000**, *298*, 163.

51 Kranz, J. K.; Hall, K. B. RNA recognition by the human U1A protein is mediated by a network of local cooperative interactions that create the optimal binding surface, *J Mol Biol* **1999**, *285*, 215.

52 Showalter, S. A.; Hall, K. B. Altering the RNA-binding mode of the U1A RBD1 protein, *J Mol Biol* **2004**, *335*, 465.

53 Luchansky, S. J.; Nolan, S. J.; Baranger, A. M. Contribution of RNA conformation to the stability of a high-affinity RNA-protein complex, *J Am Chem Soc* **2000**, *122*, 7130.

54 Shiels, J. C.; Tuite, J. B.; Nolan, S. J.; Baranger, A. M. Investigation of a conserved stacking interaction in target site recognition by the U1A protein, *Nucleic Acids Res* **2002**, *30*, 550.

55 Heinrichs, V.; Bach, M.; Winkelmann, G.; Luhrmann, R. U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site, *Science* **1990**, *247*, 69.

56 Will, C. L.; Rumpler, S.; Klein Gunnewiek, J.; van Venrooij, W. J.; Luhrmann, R. In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes, *Nucleic Acids Res* **1996**, *24*, 4614.

57 Green, M. R. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing, *Annu Rev Cell Biol* **1991**, *7*, 559.

58 Shao, W.; Kim, H. S.; Cao, Y.; Xu, Y. Z.; Query, C. C. A U1-U2 snRNP interaction network during intron definition, *Mol Cell Biol* **2012**, *32*, 470.

59 Krummel, D. A.; Nagai, K.; Oubridge, C. Structure of spliceosomal ribonucleoproteins, *F1000 Biol Rep* **2010**, *2*.

60 Colgan, D. F.; Manley, J. L. Mechanism and regulation of mRNA polyadenylation, *Genes Dev* **1997**, *11*, 2755.

61 van Gelder, C. W.; Gunderson, S. I.; Jansen, E. J.; Boelens, W. C.; Polycarpou-Schwarz, M.; Mattaj, I. W.; van Venrooij, W. J. A complex secondary structure in U1A pre-mRNA

that binds two molecules of U1A protein is required for regulation of polyadenylation, *EMBO J* **1993**, *12*, 5191.

62 Gunderson, S. I.; Beyer, K.; Martin, G.; Keller, W.; Boelens, W. C.; Mattaj, L. W. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase, *Cell* **1994**, *76*, 531.

63 Varani, L.; Gunderson, S. I.; Mattaj, I. W.; Kay, L. E.; Neuhaus, D.; Varani, G. The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein, *Nat Struct Biol* **2000**, *7*, 329.

64 Gunderson, S. I.; Vagner, S.; Polycarpou-Schwarz, M.; Mattaj, I. W. Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation, *Genes Dev* **1997**, *11*, 761.

65 Klein Gunnewiek, J. M.; Hussein, R. I.; van Aarssen, Y.; Palacios, D.; de Jong, R.; van Venrooij, W. J.; Gunderson, S. I. Fourteen residues of the U1 snRNP-specific U1A protein are required for homodimerization, cooperative RNA binding, and inhibition of polyadenylation, *Mol Cell Biol* **2000**, *20*, 2209.

66 Millevoi, S.; Vagner, S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation, *Nucleic Acids Res* **2010**, *38*, 2757.

67 Phillips, C.; Jung, S.; Gunderson, S. I. Regulation of nuclear poly(A) addition controls the expression of immunoglobulin M secretory mRNA, *EMBO J* **2001**, *20*, 6443.

68 Phillips, C.; Gunderson, S. Sequences adjacent to the 5' splice site control U1A binding upstream of the IgM heavy chain secretory poly(A) site, *J Biol Chem* **2003**, *278*, 22102.

69 Phillips, C.; Pachikara, N.; Gunderson, S. I. U1A inhibits cleavage at the immunoglobulin M heavy-chain secretory poly(A) site by binding between the two downstream GU-rich regions, *Mol Cell Biol* **2004**, *24*, 6162.

70 Ma, J.; Gunderson, S. I.; Phillips, C. Non-snRNP U1A levels decrease during mammalian B-cell differentiation and release the IgM secretory poly(A) site from repression, *RNA* **2006**, *12*, 122.

71 Milcarek, C.; Martincic, K.; Chung-Ganster, L. H.; Lutz, C. S. The snRNP-associated U1A levels change following IL-6 stimulation of human B-cells, *Mol Immunol* **2003**, *39*, 809.

72 Hall-Pogar, T.; Liang, S.; Hague, L. K.; Lutz, C. S. Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR, *RNA* **2007**, *13*, 1103.

73 O'Connor, J. P.; Alwine, J. C.; Lutz, C. S. Identification of a novel, non-snRNP protein complex containing U1A protein, *RNA* **1997**, *3*, 1444.

74 Lutz, C. S.; Alwine, J. C. Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal, *Genes Dev* **1994**, *8*, 576.

75 Liang, S.; Lutz, C. S. p54nrb is a component of the snRNP-free U1A (SF-A) complex that promotes pre-mRNA cleavage during polyadenylation, *RNA* **2006**, *12*, 111.

76 Hall-Pogar, T.; Zhang, H.; Tian, B.; Lutz, C. S. Alternative polyadenylation of cyclooxygenase-2, *Nucleic Acids Res* **2005**, *33*, 2565.

77 Lu, J. R.; Hall, K. B. An Rbd That Does Not Bind Rna - Nmr Secondary Structure Determination and Biochemical-Properties of the C-Terminal Rna-Binding Domain from the Human U1a Protein, *J Mol Biol* **1995**, *247*, 739.

78 Williams, D. J.; Hall, K. B. RNA hairpins with non-nucleotide spacers bind efficiently to the human U1A protein, *J Mol Biol* **1996**, *257*, 265.

79 Tsai, D. E.; Harper, D. S.; Keene, J. D. U1-snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts, *Nucleic Acids Res* **1991**, *19*, 4931.

80 Jovine, L.; Oubridge, C.; Avis, J. M.; Nagai, K. Two structurally different RNA molecules are bound by the spliceosomal protein U1A using the same recognition strategy, *Structure* **1996**, *4*, 621.

81 Keyes, L. N.; Cline, T. W.; Schedl, P. The primary sex determination signal of Drosophila acts at the level of transcription, *Cell* **1992**, *68*, 933.

82 Salz, H. K.; Erickson, J. W. Sex determination in Drosophila: The view from the top, *Fly (Austin)* **2010**, *4*, 60.

83 Erickson, J. W.; Quintero, J. J. Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in Drosophila, *PLoS Biol* **2007**, *5*, e332.

84 Cline, T. W. The Drosophila sex determination signal: how do flies count to two?, *Trends Genet* **1993**, *9*, 385.

85 Erickson, J. W.; Cline, T. W. A bZIP protein, sisterless-a, collaborates with bHLH transcription factors early in Drosophila development to determine sex, *Genes Dev* **1993**, *7*, 1688.

86 Jinks, T. M.; Polydorides, A. D.; Calhoun, G.; Schedl, P. The JAK/STAT signaling pathway is required for the initial choice of sexual identity in Drosophila melanogaster, *Mol Cell* **2000**, *5*, 581.

87 Sefton, L.; Timmer, J. R.; Zhang, Y.; Beranger, F.; Cline, T. W. An extracellular activator of the Drosophila JAK/STAT pathway is a sex-determination signal element, *Nature* **2000**, *405*, 970.

88 Avila, F. W.; Erickson, J. W. Drosophila JAK/STAT pathway reveals distinct initiation and reinforcement steps in early transcription of Sxl, *Curr Biol* **2007**, *17*, 643.

89 Horabin, J. I.; Schedl, P. Regulated Splicing of the Drosophila Sex-Lethal Male Exon Involves a Blockage Mechanism, *Mol Cell Biol* **1993**, *13*, 1408.

90 Horabin, J. I.; Schedl, P. Sex-lethal autoregulation requires multiple cis-acting elements upstream and downstream of the male exon and appears to depend largely on controlling the use of the male exon 5' splice site, *Mol Cell Biol* **1993**, *13*, 7734.

91 Nagengast, A. A.; Stitzinger, S. M.; Tseng, C. H.; Mount, S. M.; Salz, H. K. Sex-lethal splicing autoregulation in vivo: interactions between SEX-LETHAL, the U1 snRNP and U2AF underlie male exon skipping, *Development* **2003**, *130*, 463.

92 Siera, S. G.; Cline, T. W. Sexual back talk with evolutionary implications: stimulation of the Drosophila sex-determination gene sex-lethal by its target transformer, *Genetics* **2008**, *180*, 1963.

93 Black, D. L. Mechanisms of alternative pre-messenger RNA splicing, *Annu Rev Biochem* **2003**, *72*, 291.

94 Penalva, L. O.; Sanchez, L. RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation, *Microbiol Mol Biol Rev* **2003**, *67*, 343.

95 Camara, N.; Whitworth, C.; Van Doren, M. The creation of sexual dimorphism in the Drosophila soma, *Curr Top Dev Biol* **2008**, *83*, 65.

96 Valcarcel, J.; Singh, R.; Zamore, P. D.; Green, M. R. The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA, *Nature* **1993**, *362*, 171.

97 Boggs, R. T.; Gregor, P.; Idriss, S.; Belote, J. M.; McKeown, M. Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transformer gene, *Cell* **1987**, *50*, 739.

98 Sosnowski, B. A.; Belote, J. M.; McKeown, M. Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage, *Cell* **1989**, *58*, 449.

99 Bashaw, G. J.; Baker, B. S. The regulation of the Drosophila msl-2 gene reveals a function for Sex-lethal in translational control, *Cell* **1997**, *89*, 789.

100 Kelley, R. L.; Wang, J.; Bell, L.; Kuroda, M. I. Sex lethal controls dosage compensation in Drosophila by a non-splicing mechanism, *Nature* **1997**, *387*, 195.

101 Abaza, I.; Coll, O.; Patalano, S.; Gebauer, F. Drosophila UNR is required for translational repression of male-specific lethal 2 mRNA during regulation of X-chromosome dosage compensation, *Genes Dev* **2006**, *20*, 380.

102 Duncan, K.; Grskovic, M.; Strein, C.; Beckmann, K.; Niggeweg, R.; Abaza, I.; Gebauer, F.; Wilm, M.; Hentze, M. W. Sex-lethal imparts a sex-specific function to UNR by recruiting it to the msl-2 mRNA 3' UTR: translational repression for dosage compensation, *Genes Dev* **2006**, *20*, 368.

103 Patel, G. P.; Ma, S.; Bag, J. The autoregulatory translational control element of poly(A)-binding protein mRNA forms a heteromeric ribonucleoprotein complex, *Nucleic Acids Res* **2005**, *33*, 7074.

104 Duncan, K. E.; Strein, C.; Hentze, M. W. The SXL-UNR corepressor complex uses a PABP-mediated mechanism to inhibit ribosome recruitment to msl-2 mRNA, *Mol Cell* **2009**, *36*, 571.

105 Robida, M. D.; Rahn, A.; Singh, R. Genome-wide identification of alternatively spliced mRNA targets of specific RNA-binding proteins, *PLoS One* **2007**, *2*, e520.

106 Gawande, B.; Robida, M. D.; Rahn, A.; Singh, R. Drosophila Sex-lethal protein mediates polyadenylation switching in the female germline, *EMBO J* **2006**, *25*, 1263.

107 Penn, J. K.; Schedl, P. The master switch gene sex-lethal promotes female development by negatively regulating the N-signaling pathway, *Dev Cell* **2007**, *12*, 275.

108 Singh, R.; Valcarcel, J.; Green, M. R. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins, *Science* **1995**, *268*, 1173.

109 Lee, A. L.; Volkman, B. F.; Robertson, S. A.; Rudner, D. Z.; Barbash, D. A.; Cline, T. W.; Kanaar, R.; Rio, D. C.; Wemmer, D. E. Chemical shift mapping of the RNA-binding interface of the multiple-RBD protein sex-lethal, *Biochemistry* **1997**, *36*, 14306.

110 Crowder, S. M.; Kanaar, R.; Rio, D. C.; Alber, T. Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal, *Proc Natl Acad Sci U S A* **1999**, *96*, 4892.

111 Inoue, K.; Hoshijima, K.; Sakamoto, H.; Shimura, Y. Binding of the Drosophila sex-lethal gene product to the alternative splice site of transformer primary transcript, *Nature* **1990**, *344*, 461.

# CHAPTER 2

# Small Molecule Inhibitors of RNA Recognition Motif-RNA Interactions

## *2.1 Introduction*

RNA-binding proteins play a key role in transcription and post-transcriptional events, including alternative splicing, transport, stability, localization and translation.[1-3] Moreover RNA-binding proteins bind to small interference RNAs and microRNAs, which regulate transcription, RNA stability and translation.[3,4] Therefore, small molecule inhibitors of complexes formed between RNA-binding protein and RNA would be valuable tools to study these post-transcriptional processes and could be used to regulate gene expression at the post-transcriptional level. This chapter will describe the identification of three unrelated small molecules that destabilize two different RRM-RNA complexes by binding to the protein.

### *2.1.1 Small molecule inhibitors of protein-RNA interaction*

Small molecule inhibitors have been mostly developed to target disease-related proteins or nucleic acids for the obvious need of medicine.[5,6] For the past two decades small molecules were also recognized as excellent probes to study the function of proteins and cellular processes since they provide rapid, reversible and dose-dependent control over protein functions. In the field of chemical genetics, small molecules serve as a complementary tool to classic genetics and siRNA to understand the function of a specific protein.[7-9] Given the importance of protein-RNA interactions in a wide range of cellular processes, it is necessary to discover small molecule inhibitors of protein-RNA interactions.

Small molecules that inhibit protein-RNA interactions can be divided into two classes:

RNA binders and protein binders. Many small molecules that bind to different target RNAs and inhibit the interactions of protein-RNA complexes have been reported.[10] The most extensively studied systems are Tat protein-TAR (trans-activating response element) and Rev-RRE (Rev response element) from the human immunodeficiency virus type 1 (HIV-1). Tat protein facilitates the transcription of the HIV-1 genome by binding to TAR, which is a bulged RNA hairpin loop present in the 5' end of all mRNAs. In a complementary step, the Rev protein regulates splicing and exports the transcribed RNAs from the nucleus of the host cell to the cytoplasm. The Rev protein binds with high affinity and specificity to a stem loop within the RRE and promotes oligomerization of Rev proteins. This process triggers the "late" gene expression. Therefore, inhibition of the Tat-TAR and the Rev-RRE complexes would have a strong negative effect on the HIV life cycle. Since both TAR and RRE present unique structural features that can be targeted with small molecules, many researchers focused on finding compounds that bind to the RNA. As a result, numerous small molecules including intercalators, intercalator-arginine conjugates, guanidinylated compounds, aminoquinolones, diphenylfurans, aminoglycosides, acridines were found to inhibit these complexes.[10]

In some cases it is necessary to disrupt the protein-RNA interaction by binding to the RNA and not the protein. Myotonic dystrophy is a genetic disease caused by a trinucleotide repeat expansion.[11] In type I mytonic dystrophy (DM1), there is an aberrant expansion of CTG repeats (from 80 to >2,000 repeats) in the 3' untranslated region of the *DMPK* gene.[12] When the expanded sequence is transcribed into the RNA, it sequesters a critical splicing regulatory protein, muscleblind-like (MBNL) protein, which results in missplicing of several target pre-mRNAs. For therapeutic agents, compounds that bind the RNA and do not bind to the protein should be identified. Several compounds that specifically bind to the CUG repeat of mRNA have been

identified and some of them have been shown to reverse the splicing defects.[13-16]

Examples of small molecules that bind to the protein and inhibit protein-RNA interactions are rare. Hamy and co-workers have reported a disulfonated stilbene compound that inhibits the Tat-TAR complex.[17] They proposed this compound binds to the Tat protein based on the fact that it did not bind to the RNA in any of the traditional RNA binding experiments and pre-incubation of the compound with Tat increased the potency. The compound was able to inhibit Tat transactivation in cells and viral replication. In another example, a computer based screen identified compounds that disrupt the binding of double-stranded (ds) RNA to toll-like receptor 3 (TLR3).[18] A compound library was screened against the dsRNA-binding domain of TLR3. Experimental analysis confirmed the activity of the hits obtained from the computer screen, however, whether the compounds bind to the protein or the RNA was not experimentally shown. Small molecules that target RNA binding proteins may be used to probe the specific function or provide information of the RNA binding mechanism. Moreover if the compound can specifically target one domain in a multi-domain RNA binding protein, it may give insight into the role of that specific domain within the protein.

*2.1.2 Small molecule inhibitors of RRM-RNA complexes*

To date, there are only a handful of compounds that inhibit RRM-RNA interactions. An aminoacridine derivative was found to inhibit the complex formed between U1A and stem loop 2 RNA (SL2). [19] Fluorescence experiments showed that the compound binds to the RNA for its inhibitory activity. Interaction of the HuR protein with the AU rich element (ARE) was targeted by several groups, since HuR has been implicated in various types of cancers.[20-22] Meisner and coworkers have identified several small molecule inhibitors of HuR-ARE complex by screening

microbial and plant extracts.[20] To confirm the compound binding site, they first hypothesized several different binding models based on whether the compound binds to the HuR monomer, dimer or the RNA. Then they calculated the binding curves of each model using a mathematical program and compared the curves with the experimental data. Based on this analysis they suggested that these compounds bind to the protein to prevent the complex formation.[20] Both stem loop 2 of U1 snRNA and ARE form stable secondary structures. There are no inhibitors of complexes formed between RRMs and unstructured RNAs, although many RRMs bind to unstructured RNA targets. The flexibility of RNA and the typically broad interaction surface between RRMs and RNAs pose a significant challenge to find small molecules that inhibit the interactions between RRM and unstructured, single-stranded RNA complexes.[23]

To add to the small number of RRM-RNA inhibitors and to investigate the possibility of inhibiting a complex between a protein and an unstructured single-stranded RNA, we performed a high throughput screen using the Sex lethal-*tra* RNA complex as a model system. We identified three micromolar inhibitors of the Sex lethal-*tra* RNA complex, which bind to Sex lethal and destabilize the complex. A selectivity assay with the U1A-SL2 complex revealed that one of the compounds is a potent and selective inhibitor of U1A-SL2 interactions. Together, we demonstrate that challenging targets such as RRM and single-stranded RNA interactions may be disrupted with small molecules that bind to the protein.

## *2.2 Results and Discussion*

### *2.2.1 Selection of Protein-RNA constructs*

To identify small molecule inhibitors of RRM-RNA interactions, we developed a fluorescence anisotropy screen to monitor the increase or decrease in Sex lethal-RNA binding

upon addition of a small molecule. Sex lethal protein has been structurally characterized by X-ray crystallography bound to a portion of the *tra* pre-mRNA (Figures 2.1a and c),[24] and its binding affinity for the *tra* pre-mRNA has been studied extensively.[25-27] Therefore, we used the Sex lethal-*tra* RNA complex as a model system for identifying small molecule inhibitors of Sex lethal-RNA interactions.

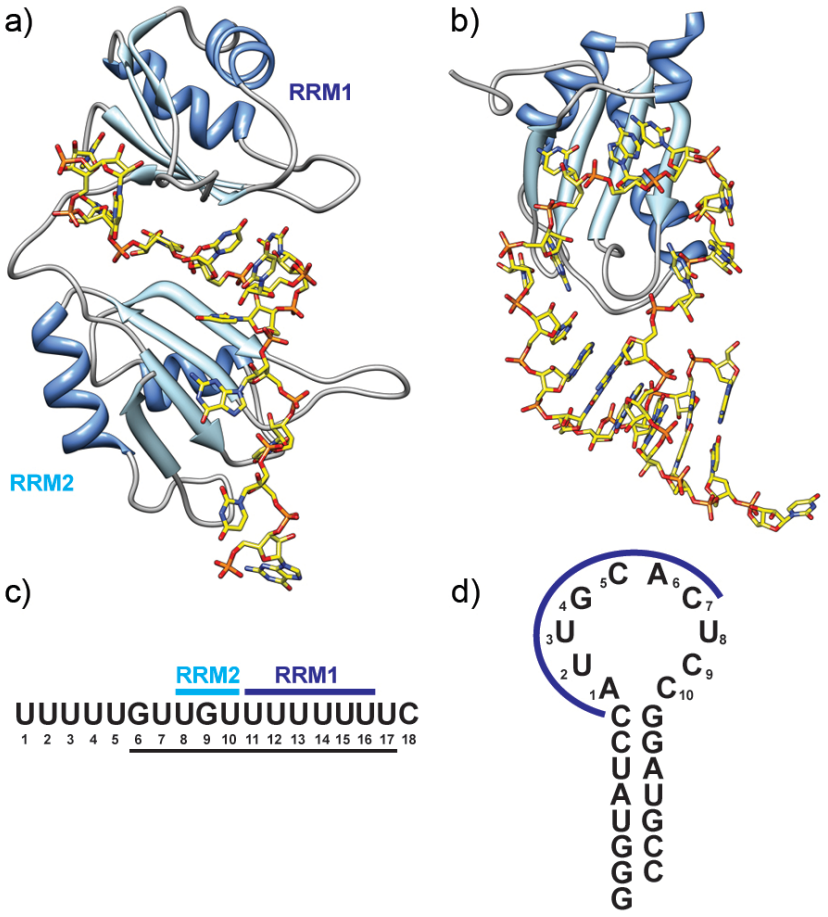



**Figure 2.1** Diagram of RRM-RNA complexes from the X-ray cocrystal structures, and RNA sequences used in this study. a) Sex lethal-*tra* RNA (pdb: 1B7F).[24] b) U1A protein-SL2 RNA complex (pdb: 1URN).[28] c) Polypyrimidine tract of *tra* RNA. The sequence shown in the structure is underlined in black. Nucleotides that have direct contact with the protein are indicated with blue lines above the sequence. d) Stem loop 2 of U1 snRNA. The 7 residues of the loop that have close contact with the protein are indicated with a blue line above the sequence. The structures were generated with the program Chimera.[29]

*2.2.2 Fluorescence anisotropy binding assay.*

For the high-throughput screen, a fluorescence anisotropy binding assay with a labeled RNA target was employed. To maximize the magnitude of the fluorescence anisotropy signal increase upon protein binding, we used a short 11mer *tra* RNA (U8~C18, Figure 1b) with a fluorescein label at the 5' end. Titration of Sex lethal protein into solutions of *tra* RNA resulted in a reproducible binding curve showing increase in fluorescence anisotropy upon complex formation (Figure 2.2). The calculated $K_d$ was comparable to that obtained from electrophoretic mobility shift assays (Figure 2.7b).

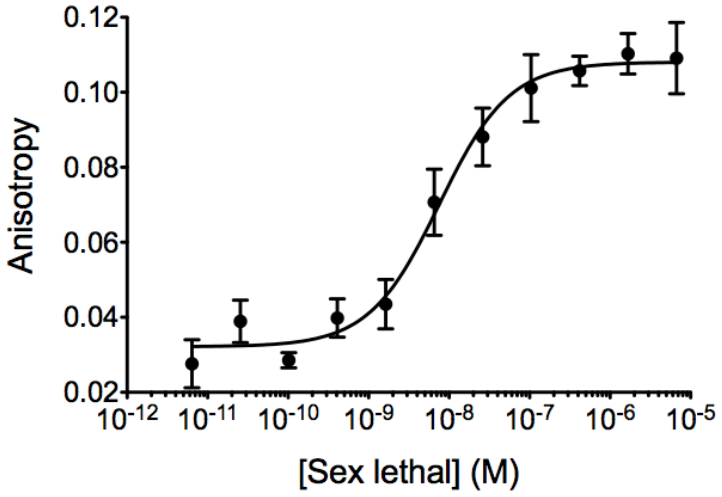



**Figure 2.2** Fluorescence anisotropy based assay for Sex lethal-*tra* RNA binding. Increasing concentration of Sex lethal protein increases the fluorescence anisotropy indicating Sex lethal protein-*tra* RNA complex formation. $K_d$ value was calculated to be 7.5 ± 1.5 nM. The data represent average of 3 independent measurements.

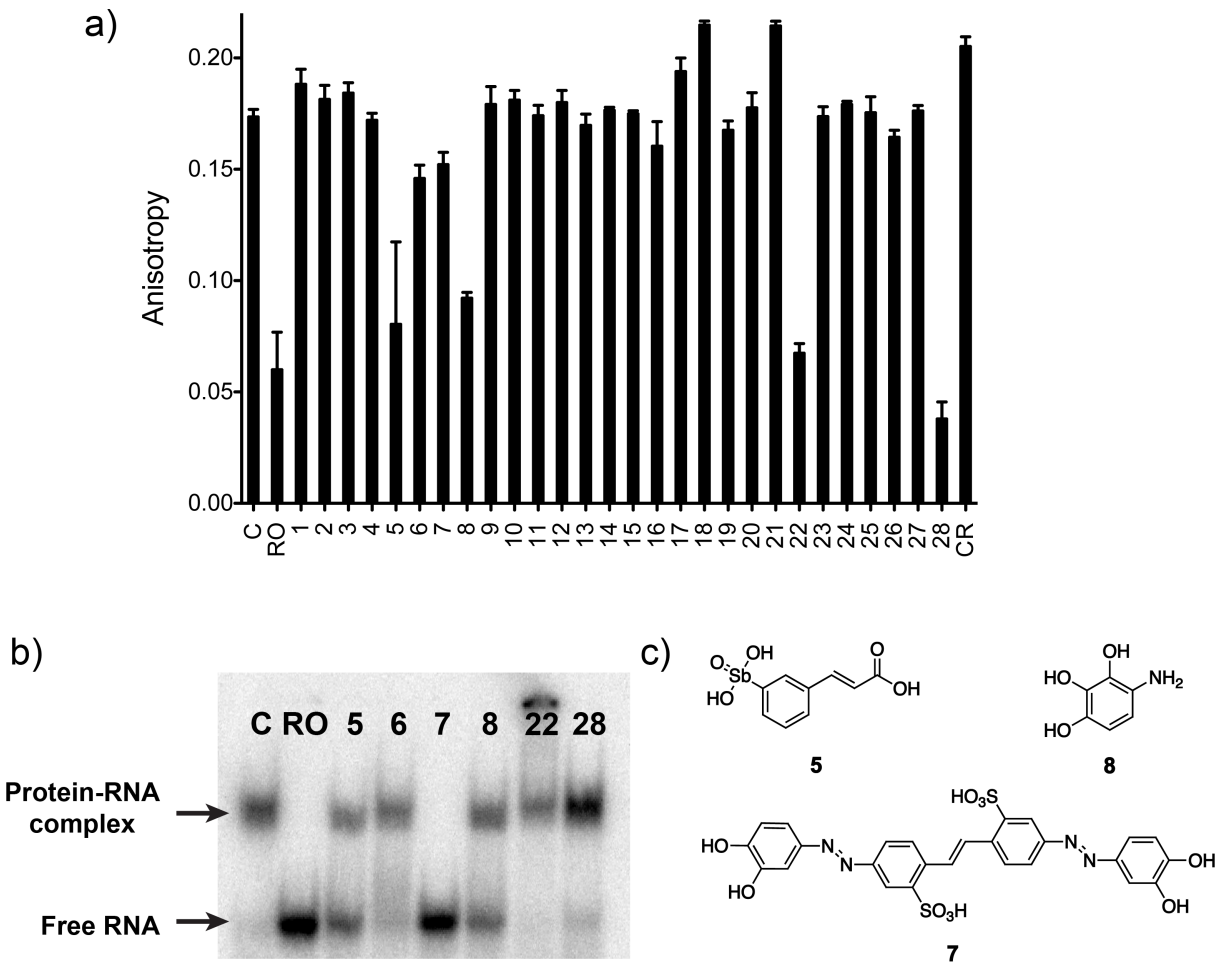



**Figure 2.3** Identification of inhibitors of the Sex lethal-*tra* RNA complex. a) Fluorescence anisotropy assay of 28 initial hits in the presence of 0.1% Triton X-100. 100 μM of each compound was incubated for 1h with 0.5 μM Sex lethal and 3 nM *tra* RNA in HEPES buffer. C is the DMSO control, RO is RNA only and CR is Congo red, a known aggregator. Error bars represent SD from 3 independent experiments. b) Six compounds (**5**, **6**, **7**, **8**, **22** and **28**) selected from the fluorescence anisotropy assay were subjected to an electrophoretic mobility shift assay in the presence of 0.5% Triton X-100 and 0.2 mg/mL BSA. 100 μM of each compound was incubated for 2h with 0.1 μM protein and 400 pM RNA. C is the DMSO control and RO is RNA only. Representative gel of three independent experiments is shown. c) Structure of compounds **5**, **7**, and **8**.

*2.2.3 Compound screening.*

We chose to utilize the Chembridge library (32,000 compounds) and NCI library (10,560 compounds) for high throughput screening because they contain compounds with a high degree of molecular structural diversity and compounds can be obtained from the NCI or purchased for experimental evaluation. Screening was carried out in 384 well plates at the high-throughput screening facility at University of Illinois. Compounds to be evaluated were added to a solution of the Sex lethal-*tra* RNA complex and the decrease in anisotropy was monitored to identify inhibitors. DMSO alone was used as the negative control, and because there is no known compound that inhibits the Sex lethal protein-RNA complex, labeled RNA alone was used as the positive control in each plate. These experiments identified 28 compounds with more than 40% inhibitory activity, which were selected as primary hits. Compounds identified by high-throughput screening are often false positives due to their ability to form aggregates in buffered-solutions.[30,31] Small molecule aggregates can isolate the protein from the RNA, which reduces the fluorescence anisotropy signal. It has been shown that 0.1% of Triton X-100 can efficiently minimize small molecule aggregation.[31] Therefore the primary hits were screened again in the presence of 0.1% Triton X-100 (Figure 2.3a). Six compounds that had lower anisotropy values than the DMSO control were used for further analysis.

*2.2.4 Electrophoretic mobility shift assay*

Because fluorescence anisotropy is a fluorescence-based method, compounds that absorb or are fluorescent at the excitation or emission wavelength can also produce false positive signals.[32] To eliminate these compounds we performed a secondary assay that does not involve a fluorescence-based readout, the electrophoretic mobility shift assay (EMSA). EMSA is based on

the observation that the migration of the protein-RNA complex is slower than the unbound radiolabeled RNA in a native polyacrylamide gel during electrophoresis.[33] The inhibition of the Sex lethal-*tra* RNA complex, was monitored by observing the decrease in the intensity of the protein-RNA complex band and the increase in the free RNA band in the gel upon addition of the compounds (Figure 2.3b). In addition to Triton X-100, bovine serum albumin (BSA) was found to attenuate compound aggregation when pre-incubated with the compound.[34] Therefore, the compounds (100 μM) were pre-incubated with 0.2 mg/mL of BSA for 10 minutes followed by the addition of the Sex lethal-*tra* RNA complex. For the lanes with compounds **6**, **22**, and **28**, less than 40% inhibition were observed. These compounds are thought to be false positives due to aggregation or fluorescence of the compounds and were discarded.

Three distinct compounds inhibited the Sex lethal-*tra* RNA complex at high concentrations of Triton X-100 and BSA (Figure 2.3b). Compounds **5**, **7**, and **8** effectively inhibited the pre-formed Sex lethal-*tra* RNA complex with 62, 93 and 40% inhibition, respectively. The structures of the three hit compounds are shown in Figure 2c. Compound **5** is an arylstibonic acid that has been found to inhibit several DNA binding proteins including bZIP transcription factor, poxyvirus type 1 topoisomerase, and human apurinic/apyrimidinic endonuclease 1.[35-37] Compound **7** is a disulfonated stilbene dye with two catechol moieties. Despite its small size, compound **8** was also identified as an inhibitor of the Sex lethal-*tra* complex.

*2.2.5 Controlling for Aggregation*

A few dyes, including Congo red, have been reported to form aggregates and promiscuously inhibit enzymes.[38] Concerned with the fact that compound **7** is a dye with azo

moieties and compound **5** has been identified as an inhibitor of other proteins we evaluated these compounds with dynamic light scattering to further explore the possibility of these compounds being aggregation-based inhibitors. Colloid-like large particles formed by small molecules can be detected with dynamic light scattering. There was no evidence that compounds **5** and **8** formed any particles, even at concentrations of 500 μM. For compound **7**, peaks appeared randomly, which were averaged to a small particle size and a high standard deviation (Table 2.1). We analyzed the autocorrelation functions of the dynamic light scattering data.[30] In comparison with the known aggregator Congo red, the autocorrelation function of compound **7** did not show a well-defined autocorrelation function. This suggests that pre-formed aggregates of compound **7** are not responsible for the observed inhibition of Sex lethal-*tra* RNA complex (Figure 2.4).

**Table 2.1** Dynamic light scattering data of Congo red and three inhibitors.

| | Congo red | **5** | **7** | **8** |
|---|---|---|---|---|
| Diameter (nm) | 210 (±28) | No particle | 6 (±11)[a] | No particle |
| Intensity (kcps) | 4.3 | 0.5 | 0.6 | 0.3 |

[a] Peaks were observed randomly reflected in the high standard deviation.

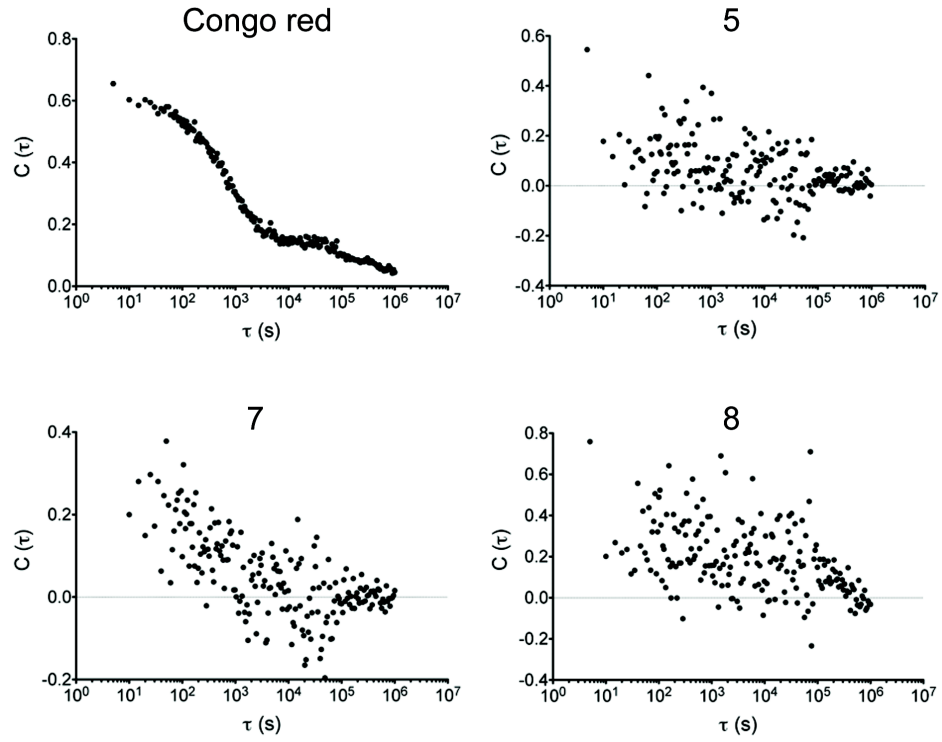



**Figure 2.4** Representative autocorrelation functions from dynamic light scattering. Compound aggregation was analyzed by dynamic light scattering. Formation of particles was monitored with 500 μM of each compound in HEPES buffer.

*2.2.6 $IC_{50}$ values.*

Dose-dependent inhibition of the Sex lethal-*tra* RNA complex by the selected compounds was analyzed using EMSA. The concentration of Sex lethal was adjusted so that 80% of the radiolabeled *tra* RNA was bound to the protein. To this solution of pre-formed complex, increasing amounts of compounds were added and subjected to electrophoresis (Figure 2.5a). We found that compound **5**, **7,** and **8** inhibit the Sex lethal-*tra* RNA complex dose-dependently with $IC_{50}$ values of 22 ± 1 μM, 5 ± 1 μM and 14.4 ± 0.4 μM, respectively (Table 2.2).

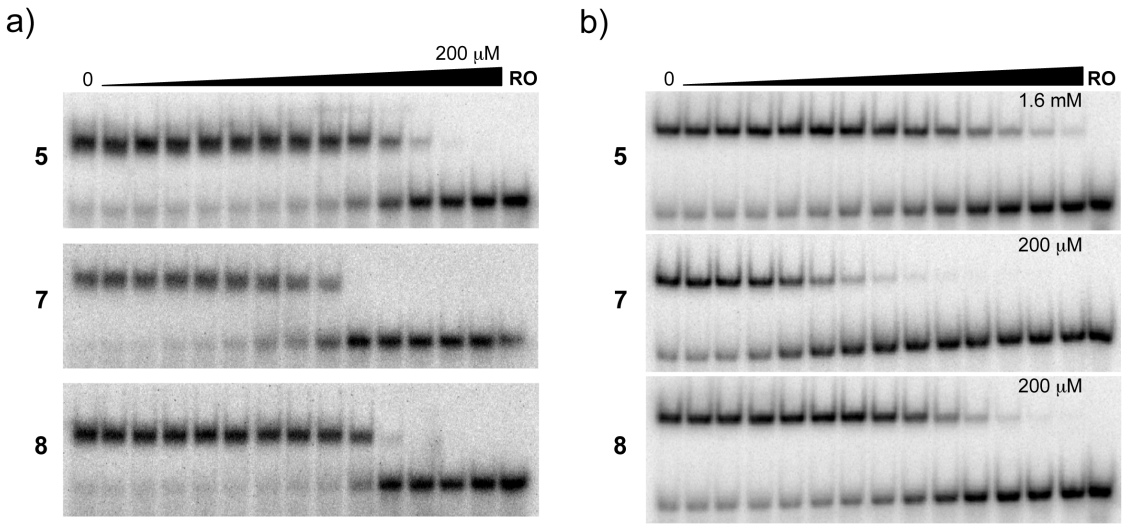



**Figure 2.5** Representative EMSA gels showing dose-dependent titrations of compounds into solutions of the protein-RNA complex (~80% fraction bound). a) Sex lethal-*tra* RNA complex. Highest concentrations of the compounds were 200 μM. Compounds were incubated with 0.1 μM Sex lethal and 400 pM *tra* RNA in HEPES buffer that contains 0.5 % Triton X-100. b) U1A-SL2 complex. Highest concentrations of the compounds are indicated on top of each gel. Compounds were incubated with 5 nM Sex lethal and 50 pM SL2 RNA in Tris buffer that contains 0.5 % Triton X-100.

**Table 2.2** $IC_{50}$ values of the three inhibitors against Sex lethal-*tra* and U1A-SL2 RNA complexes.

| **$IC_{50}$ (μM)** | **5** | **7** | **8** |
|---|---|---|---|
| **Sex lethal** | **22 (±1)** | 5 (±1) | 14.4 (±0.4) |
| **U1A** | 92 (±1) | **0.48 (±0.05)** | **9 (±1)** |

*2.2.7 Fluorescence binding assays*

To understand whether the compounds are binding to the RNA or the protein to achieve their inhibitory activity, we used fluorescence spectroscopy. First, to detect whether the compounds bind to the RNA, compounds were titrated into a 100 nM of 5'-fluorescein labeled 11mer *tra* RNA solution. We used a shorter RNA construct so that if the compounds bind to the

RNA, the fluorescein is closer to the bound site and more likely to give a signal change. Although this experiment cannot conclusively tell us whether the compounds are binding to the RNA, if we do observe a signal change we can positively identify compounds that bind to the RNA. However no fluorescein quenching was observed up to concentrations of 127 μM of the compounds (Figure 2.6). The compounds may not be binding to the RNA or they could bind to the RNA in a manner that does not quench the fluorescein. Compound **7** was not subjected to the RNA fluorescence experiment due to its large absorption at the excitation wavelength of fluorescein.

In a separate experiment, we introduced a tryptophan residue into Sex lethal as a fluorescence label. Phe256 is a highly conserved residue and participates in stacking interactions with uracil 7 of *tra* RNA.[24] This interaction is important for the binding.[39] Assuming the compounds would bind to the protein surface where it interacts with RNA to inhibit the protein-RNA interaction we mutated this residue to a tryptophan to monitor the fluorescence quenching of the Trp by the compounds. There are no other Trp residues in the protein. Before using the Sex lethal (F256W) mutant we confirmed that this phenylalanine to tryptophan mutation does not affect the overall structure of the protein by circular dichroism (CD) analysis and that the WT and the mutant proteins have comparable $K_d$ values by EMSA (Figure 2.7).

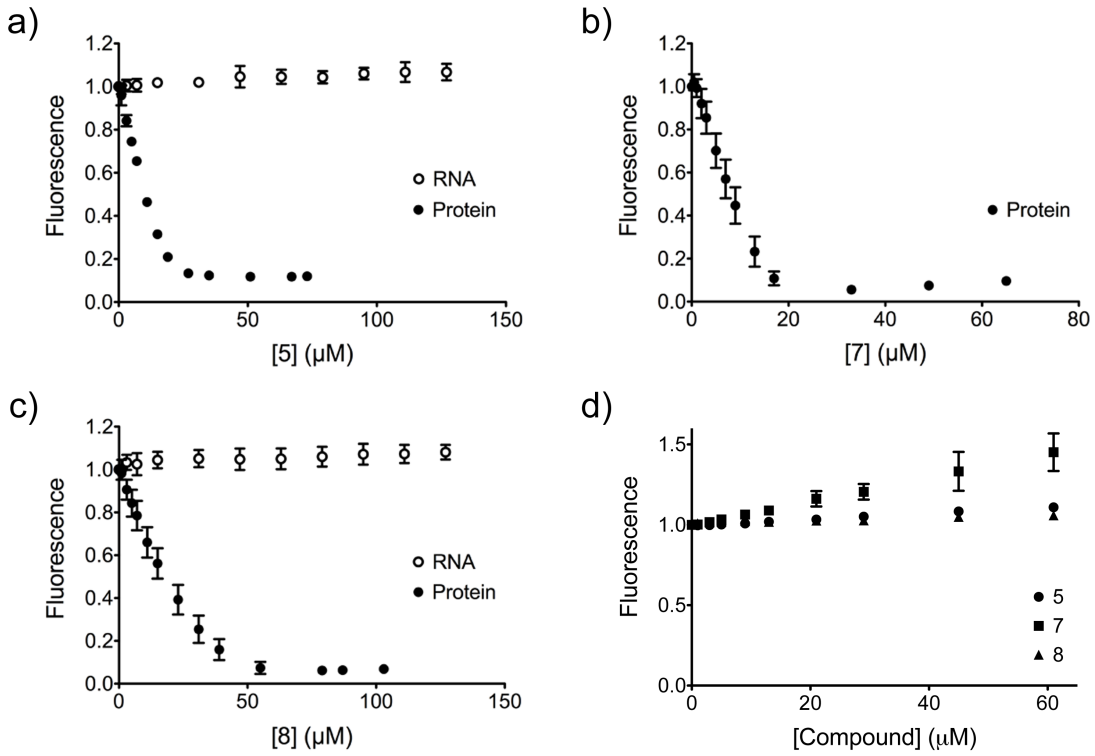



**Figure 2.6** Fluorescence-binding assays indicating that all three compounds bind to the protein. Two independent fluorescence-binding assays were performed using fluorescein labeled RNA and Sex lethal (F256W). Compounds were titrated into a solution of fluorescein labeled RNA or Sex lethal (F256W). Normalized fluorescence intensity of the RNA and the protein is shown in the y-axis upon titration of a) compound 5, b) compound 7 or c) compound 8. d) A control experiment with Tryptophan. Compounds were titrated into a solution of Tryptophan and the fluorecence intensity was monitored.

a)

| CD analysis | WT | F256W |
|---|---|---|
| α-helix | 24% | 26% |
| β-sheet | 40% | 39% |
| Random | 36% | 35% |

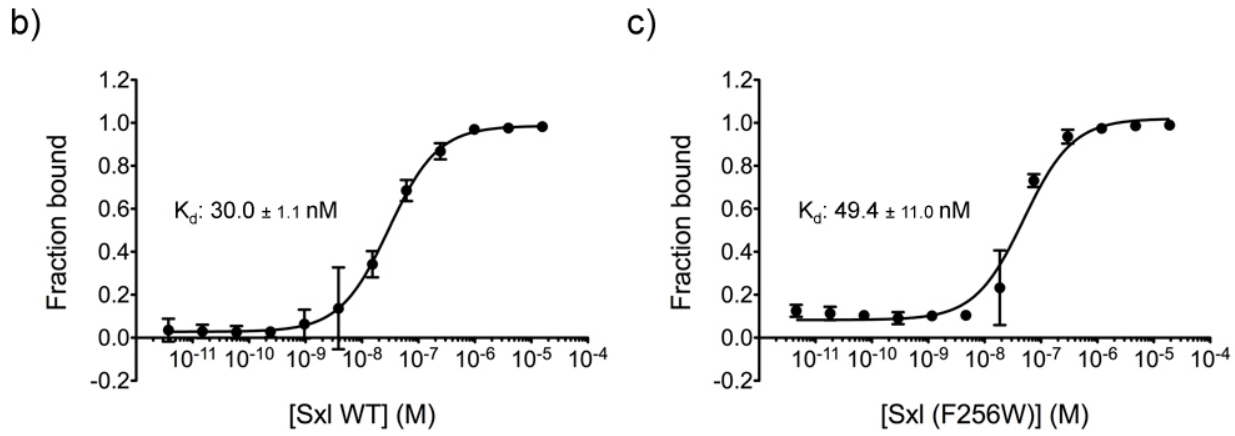



**Figure 2.7** The Phe256Trp mutation does not affect the overall structure and the binding affinity of the protein. a) Circular dichroism analysis of Sex lethal WT and Sex lethal (F256W). b) Binding curves obtained from electromobility shift assay (EMSA) of Sex lethal WT and Sex lethal (F256W).

Concentrated solutions of compounds in DMSO were titrated into a 2 μM Sex lethal solution and the Trp fluorescence was monitored at 355 nm. The change of the fluorescence signal of the mutant protein due to the addition of DMSO was subtracted from that observed upon addition of the compounds and the inner filter effect of the compounds were corrected as described in the experimental section. Normalized fluorescence intensity versus compound concentration is plotted in Figure 5. All three compounds quenched the tryptophan fluorescence of the protein dose-dependently. As a control, the fluorescence intensity of a solution of tryptophan was measured upon compound titration (Figure 2.6d), and no quenching of the tryptophan signal was observed. This result shows that the compounds do not directly bind to the free tryptophan and suggests that the fluorescence quenching observed with the F256W

mutant is a result of the compounds binding to the protein. Binding curves are shown in Figure S3. The apparent $K_d$ of the three compounds bound to Sex lethal was comparable to the $K_i$ values calculated from $IC_{50}$ values (Figure 2.6d). These results suggest that binding of the compounds to the protein prevents Sex lethal-*tra* RNA complex formation.

*2.2.8 Selectivity of the compounds*

To assess the selectivity of these compounds for the Sex lethal-*tra* RNA complex, we tested their ability to destabilize a different RRM-RNA complex, that of the U1A protein binding to SL2 RNA. We used the N-terminal RRM of U1A protein and stem loop 2 RNA to test the inhibitory activity of the compounds by EMSA (Figure 2.8b). For a fair comparison of the compounds' activity against Sex lethal and U1A, we adjusted the amount of U1A so that 80% of SL2 RNA is bound to the U1A protein. Interestingly, we observed complete inhibition of the U1A-SL2 complex at low micromolar concentration with compound **8** and high nanomolar concentration with compound **7,** whereas compound **5** did not completely inhibit complex formation even at 1.6 mM. Compound **7** has 10-fold higher potency against the U1A-SL2 RNA complex than the Sex lethal-*tra* RNA complex. Compound **8** is approximately a 2-fold stronger inhibitor of the U1A-SL2 complex compared to the Sex lethal-*tra*- RNA complex. Thus, we have identified compound **5** as a selective inhibitor of the Sex lethal-*tra* complex and compound **7** as a selective inhibitor of the U1A-SL2 RNA complex.

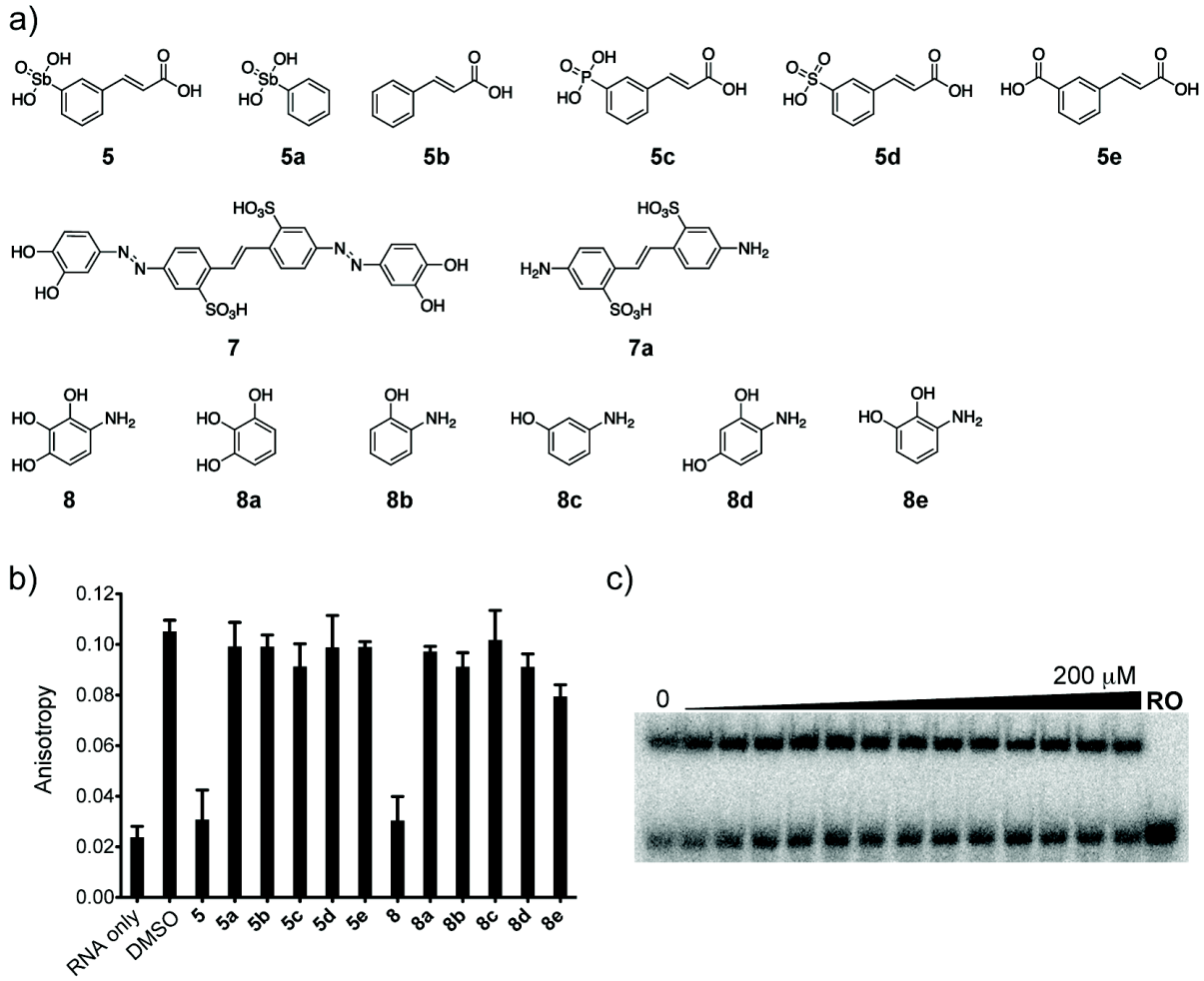



**Figure 2.8** Preliminary structure-activity relationship analysis. a) Structure of hit compounds and analogues. b) Fluorescence anisotropy was used to evaluate the ability of analogs of compounds **5** and **8** to destabilize the Sex lethal-*tra* RNA complex. c) EMSA was used to evaluate the ability of compound **7a** to destabilize the U1A-SL2 RNA complex.

*2.2.9 Structure-Activity Relationships*

A preliminary structure activity relationship analysis was performed. The core of compound **7** is a disulfonated stilbene, which could possibly mimic a diphosphate of RNA. Hamy and coworkers identified a similar disulfonated stilbene compound (**CGA127053**) as an inhibitor of Tat protein-TAR RNA complex (Figure 2.9).[17] They also found that **Ba12330**,

which does not have the acetoacetate group, had significantly lower activity. We evaluated an analog (**7a**), which does not have catechol moieties, against U1A-SL2 RNA complex (Figure 2.6). Similar to the finding in the report of Hamy et al. the core molecule itself does not inhibit the U1A-SL2 RNA complex. This suggests that the core moiety does not promiscuously bind to any nucleotide binding protein and the auxiliary functional groups are necessary for the binding affinity and may determine the selectivity of the compound.

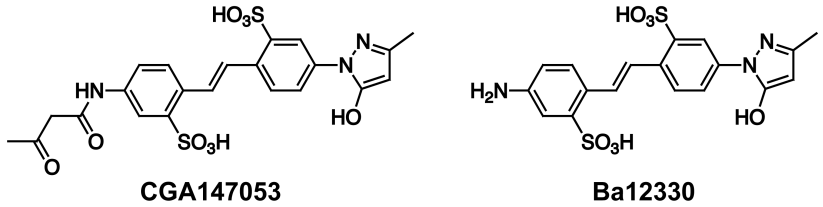



**Figure 2.9** Sulfonated stilbene compounds identified by Hamy *et al.* that are active (GCA147053) and inactive (Ba12330) against the Tat-TAR RNA complex.[17]

Arylstibonic acids have been reported to inhibit several DNA binding proteins.[40,41] The stibonic acid group of compound **5** has been thought to mimic a phosphate group for its activity against apurinic/apyrimidinic endonuclease 1.[37] To determine that possibility, we synthesized a phosphonic acid analog (**5c**), a sulfonic acid analog (**5d**), and a carboxylic acid analog (**5e**) of compound **5** and evaluated these compounds as inhibitors of the Sex lethal-*tra* RNA complex. However, none of the compounds destabilize the complex. This result suggests that the larger size of antimony and the ionic state of the stibonic acid may be critical for inhibitory activity. The results obtained with compounds **5a** and **5b** indicate that both stibonic acid and the acrylate moiety are required for activity.

Several analogs of compound **8** were also evaluated. Compound **8e** at 100 μM concentration results in 32% inhibition of Sex lethal-*tra* RNA complex, while the other analogs

(**8a**, **8b**, **8c**, and **8d**) showed negligible inhibition. Therefore all four groups of the benzene ring are important for its optimal inhibitory activity.

### *2.3 Conclusion*

In conclusion, we have identified three distinct compounds, compounds **5**, **7**, and **8** that inhibit RRM-RNA interactions. Compound **5** selectively inhibits the Sex lethal-*tra* RNA complex and is the first example of an inhibitor of a complex formed between an unstructured single-stranded RNA and a protein. Compound **7** selectively inhibits the U1A-SL2 RNA complex. All three compounds bind to the protein, instead of the RNA. Considering the fact that RRMs bind to single-stranded RNA and most of the RNA targets are unstructured, targeting the protein could be a promising approach for disrupting RRM-RNA interactions. Binding small molecules to flexible, unstructured sequences of RNA is a significant challenge. Furthermore, compounds that can specifically bind to different RRMs may be used to probe the role of each RRMs within the same protein. In fact, Meisner and coworkers have identified compounds that specifically bind to RRM3 of HuR protein and used those compounds to assign the specific function of RRM3.[22] Further analysis of the ability of these three compounds and analogs to destabilize different complexes formed between RNA and RRM-containing proteins may give insight into which functional groups on the molecules are important for specific binding. These compounds provide a basis for the design of new probes for RRM-containing proteins.

## *2.4 Materials and Methods*

### Materials

Compounds for the initial screen were obtained from the high-throughput screening facility at University of Illinois at Urbana-Champaign. 28 compounds for the second screen were obtained from NCI and purchased from Hit2Lead and confirmed with mass spectrometry (ESI). Analogs **5c**, **5d** and **8e** were synthesized and the other analogs were purchased from Sigma-Aldrich and TCI America. All RNAs were purchased from Integrated DNA technology (IDT) and purified using denaturing-PAGE.

### Protein expression and purification

His6-containing plasmids were transformed into *Escherichia coli* strain BL21DE3 (pLysS) competent cells. The cells were grown in LB medium, and protein expression was induced with 1 mM IPTG at $OD_{600}$=0.60. The cultures were grown for 5–6 h after induction. The cells were pelleted, resuspended in 10 mL of lysis buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 20 mM imidazole, pH 7.6) and lysed by ultrasonication. The lysate was centrifuged at 10,000 rpm and the supernatant was loaded on a 1 mL Ni-NTA column, and the protein was eluted with elution buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 250 mM imidazole, pH 7.6). Eluted protein was dialyzed against storage buffer (Sex lethal: 20 mM HEPES, 25 mM KCl, 1 mM β-mercaptoethanol, 1 mM EDTA; U1A: 10 mM potassium phosphate, 50 mM KCl) and concentrated by amicon filter MWCO 3,000. The protein concentration was determined by UV spectroscopy and the extinction coefficient obtained by amino acid analysis (Yale Keck facility). The protein identity was confirmed by low-resolution electrospray ionization mass spectrometry and the purities of the proteins were assessed by SDS-PAGE. Sex lethal (F256W) was made by

site-directed mutagenesis and was expressed and confirmed as the WT protein. The concentration of the Sex lethal (F256W) was determined using BCA assay (Pierce).

**Fluorescence anisotropy**

A 5'-fluorescein labeled 11mer *tra* RNA (5'-UGUUGUUUUUUUU-3') was used for the fluorescence anisotropy experiments. To assess the binding affinity of Sex lethal protein-tra RNA complex, a varying amount of Sex lethal was incubated with 3 nM *tra* RNA in a buffer solution of 15 mM HEPES, pH 7.4, 50 mM KCl, 1 mM EDTA, 1 mM , β-mercaptoethanol and 0.01% Triton X-100. The raw anisotropy values at different concentration of Sex lethal protein were fit to equation **1** to obtain the $K_d$ value.

$$y = \max \frac{1}{1 + \frac{K_d}{x}} \qquad \textbf{Eq. 1}$$

For the high-throughput screen, the concentrations of Sex lethal and 11-mer *tra* RNA were adjusted so that 80% binding was achieved. Final concentration of Sex lethal and *tra* RNA were 0.5 μM and 3 nM in a buffer solution of 15 mM HEPES, pH 7.4, 50 mM KCl, 1 mM EDTA, 1 mM , β-mercaptoethanol and 0.01% Triton X-100. 2 μL of compound solutions in DMSO (1 mM) were added to 38 μL of complex solutions in a 384 well plate. For the negative control 2 μL of DMSO was added. A RNA only sample was used as a positive control. The solution was incubated at room temperature for 1 h and the fluorescence anisotropy was measured using Analyst HT (Molecular Devices) with an excitation filter of 485 ± 10 nm and an emission filter of 530 ± 10 nm. The anisotropy was calculated by equation **2**, where r is the

anisotropy value, G is the G factor, $I_{//}$ is the fluorescence intensity with parallel-oriented and $I_{\perp}$ is the fluorescence intensity with perpendicular-oriented excitation and emission polarizers.

$$r = \frac{(I_{//} - I_{\perp})}{(I_{//} + 2GI_{\perp})} \qquad \textbf{Eq. 2}$$

The percent inhibition was calculated with equation 3, where *r* is the anisotropy value of the sample, RO is RNA only and NC is the negative control.

$$\left(1 - \frac{r - RO}{NC - RO}\right) \times 100 \qquad \textbf{Eq. 3}$$

The second screen was performed using the same condition except with a higher concentration of Triton X-100 (0.1%). The second screen was repeated three times. Analogs of the compound **5** and compound **8** were tested in the presence of 0.01% Triton X-100, 100 μM compound, and RNA 3 nM.

**Electrophoresis mobility shift assay (EMSA)**

*Sex lethal-tra RNA*. 0.1 μM of protein was incubated with 400 pM $^{32}$P-labeled RNA and varying amounts of compounds for 2 hours at 4 °C in a buffer containing 15 mM HEPES, pH 7.4, 50 mM KCl, 1 mM EDTA, 1 mM β-mercaptoethanol, 20 % glycerol, and 0.01~0.5 % Triton X-100. Electrophoreses were carried out at 4 °C for 45 minutes at 320 V with 0.5X TBE as running buffer. The gels were pre-run at 320 V for 30 minutes before loading 10 μl of reaction mixture per well. The native gels, 6% acrylamide 80:1 (acrylamide : bisacrylamide), were poured twenty-four hours in advance or a minimum of two hours before running and left to equilibrate in a cold room at 4 °C.

*U1A-SL2*. 5 nM of protein was incubated with 50 pM $^{32}$P-labeled RNA and varying amounts of compounds for 40 minutes at room temperature in a buffer containing 10 mM Tris pH 7.4, 250 mM NaCl, 1 mM EDTA and 0.5% Triton X-100. Electrophoreses were carried out at 25 °C for 30 minutes at 350 V with 1X TBE as running buffer. The native gels, 8% acrylamide 42:1 (acrylamide : biacrylamide), were pre-run at 350 V for 30 minutes before loading 10 μl of reaction mixture per well.

The gels were exposed by autoradiography (Molecular Dynamics) and analyzed with Quantity One (Biorad) software. The gels were repeated three times for each compound. The data were fit to equation **4**, using Prism v5.0 to obtain $IC_{50}$ values. *x* is the concentration of the compound added, h is the hill slope, max is the maximum binding without the compound, and min is the minimum binding.

$$y = \min + \left( \frac{\max - \min}{1 + 10^{(x - \log IC_{50})h}} \right) \qquad \textbf{Eq. 4}$$

Inhibitory constant $K_i$ was calculated using equation **5**, where $[P]_T$ is the concentration of the protein.

$$K_i = \frac{IC_{50}}{([P]_T / K_d) - 1} \qquad \textbf{Eq. 5}$$

**Fluorescent binding experiments**

All fluorescence data was obtained using a temperature controlled Horiba Jobin Yvon fluorimeter with a 250 μL cell with a 3 mm path. The buffer used was 15 mM HEPES, pH 7.4, 50 mM KCl, 1 mM EDTA, and 1 mM β-mercaptoethanol.

*RNA binding experiment.* 5'-fluorescein labeled 11mer *tra* RNA was used for the RNA binding experiments. To a 200 μL solution of 100 nM tra RNA, 1 μL of concentrated samples of compounds (200 μM, 400 μM, 800 μM, 1.6 mM and 3.2 mM) in DMSO was added for each titration. After 1 min, the fluorescence was measured with an excitation wavelength of 485 nm (slit width of 5 nm) and emission wavelength of 520 nm (slit width of 5 nm). A control experiment was performed by adding DMSO to the free RNA solution.

*Protein binding experiment.* For the binding assay with Sex lethal protein, a cell sample (200 μL) of 2 μM Sex lethal (F256W) was excited at 295 nm (slit width of 4 nm) and emission recorded between 310-450 nm (slit width of 8 nm), monitoring the maximum at 355 nm. 1 μL of concentrated samples of compounds (200 μM, 400 μM, 800 μM, 1.6 mM and 3.2 mM) in DMSO was added for each titration. The sample was then allowed to equilibrate for 1 min and the fluorescence was measured. The decrease in fluorescence with increasing concentration of compound was monitored until the fluorescence change was no longer observed. Innerfilter effect of the small molecules were calculated based on UV absorption measurements between 240~600 nm of the compounds on a Shimadzu UV2450 spectrophotometer. A control experiment was performed by adding DMSO to the protein solution. The control fluorescence intensity was subtracted from the raw data and the innerfilter effect was corrected using equation **6**.

$$F_{corr} = F_{obs} \cdot 10^{\left(\frac{OD_{ex}+OD_{em}}{2}\right)} \qquad \textbf{Eq. 6}$$

The innerfilter effect corrected fluorescence data were fit to equation **7**, using Graphpad Prism v5.0 to obtain apparent equilibrium dissociation constants ($K_d$). F is the fluorescence intensity of

the sample, $F_0$ is the initial fluorescence intensity, $F_f$ is the final fluorescence intensity, $x$ is the concentration of compound added, and h is the hill slope.

$$\frac{F - F_0}{F_f - F_0} = \max \frac{x^h}{K_d{}^h + x} \qquad \textbf{Eq. 7}$$

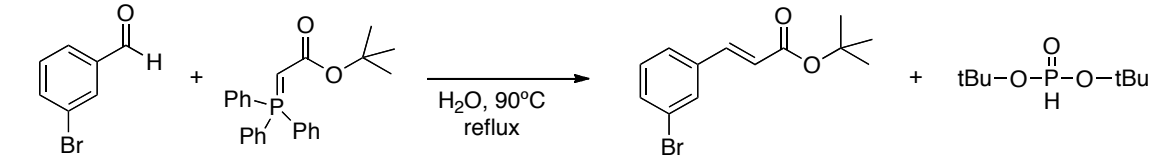
*Tryptophan control.* To a 200 μL solution of 2 μM Tryptophan, 1 μL of concentrated samples of compounds (200 μM, 400 μM, 800 μM, 1.6 mM and 3.2 mM) in DMSO was added. The measurements were performed as the protein binding experiment. The plot is an average of two independent measurements.

**Dynamic light scattering (DLS)**

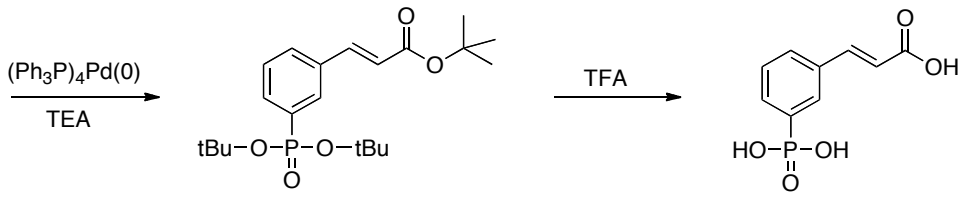
Compounds were dissolved to 20 mM in DMSO and diluted with filtered 15 mM HEPES buffer (pH 7.4) to a final concentration of 500 μM. Solutions of the compounds were analyzed using ZetaPals instrument (Brookhaven Instruments Corporation) equipped with helium-neon laser at 658 nm. Scattered light is collected at an angle of 90° by a photon-counting photomultiplier tube and is then directed to a correlator. The software (BIC particle sizing) derives particle sizes from the correlator function. Each compound was measured three times or more at room temperature.

**Synthesis of 5c, 5d and 8e**

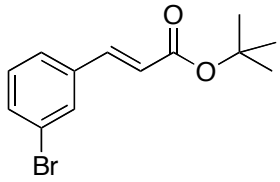
**(*E*)-3-(3-phosphonophenyl)acrylic acid (5c)**




**5c-1**



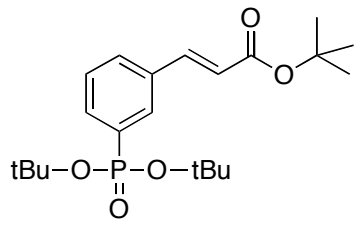

**5c-2** **5c**




**(*E*)-tert-butyl-3-(3-bromophenyl)acrylate (5c-1).**

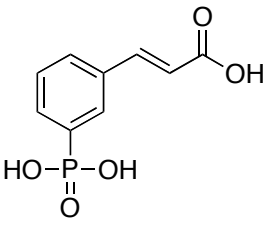
**5c-1** was obtained from a Wittig reaction between 3-bromobenzaldehyde and (tert-butoxycarbonylmethylene) triphenylphosphorane.[42] 3-bromobenzaldehyde (1.0 mmol, 185 mg) and (tert-butoxycarbonylmethylene) triphenylphosphorane (1.5 mmol, 565 mg) were stirred in water at 90°C for 2 h. After cooling the reaction mixture to room temperature, the product was extracted from the water layer with $CH_2Cl_2$ (3 × 5 mL) and the combined organic layers were dried over sodium sulfate, filtered and concentrated under vacuum. The crude product was purified using flash column chromatography (5% ETOAc/hexane) to give 55% (155 mg) of **5c-1** (E/Z ratio: 92/8). $^1$H NMR (400 MHz, $CDCl_3$) δ 7.63 (t, aromatic, 1H), 7.47 (d, C$H^{\beta}$=CH$^{\alpha}$,

J=16.0 Hz, 1H), 7.47-7.22 (m, aromatic, 3H), 6.34 (d, $CH^{\beta}$=C$H^{\alpha}$, J=16.0 Hz, 1H), 1.53 (s, ($CH_3$)$_3$CO, 9H).




**(*E*)-tert-butyl 3-(3-(di-tert-butoxyphosphoryl)phenyl)acrylate (5c-2).**

**5c-1** and di-*tert*-butylphosphite were coupled using $Pd(PPh_3)_4$.[43] In a disposable tube with screw cap was placed **5c-1** (0.55 mmol, 155 mg), di-*tert*-butylphosphite (0.66 mmol, 128 mg), $Pd(PPh_3)_4$ (0.027 mmol, 31.7 mg) and 250 μL of dry TEA/toluene (1:1). The reaction mixture was stirred under nitrogen at 80°C overnight. White precipitate was formed 1 h after the reaction mixture reached 80°C. The reaction mixture was cooled to room temperature and filtered. The filtrate was concentrated under vacuum. The residue was purified by flash column chromatography on silica gel, eluting with ETOAc/hexane (1:1) yielding 43.2 mg (20%) of **5c-2**. $^1$H NMR (500 MHz, $CDCl_3$) δ 7.91 (d, aromatic, 1H), 7.76 (m, aromatic, 1H), 7.60 (d, aromatic, 1H), 7.60 (d, C$H^{\beta}$=C$H^{\alpha}$, J=16.0 Hz, 1H), 7.42 (m, aromatic, 1H), 6.42 (d, $CH^{\beta}$=C$H^{\alpha}$, J=16.0 Hz, 1H), 1.53 (s, ($CH_3$)$_3$CO, 9H), 1.46 (s, 2×($CH_3$)$_3$PO, 18H).




**(*E*)-3-(3-phosphonophenyl)acrylic acid (5c)**

**5c-2** (0.11 mmol, 43.2 mg) was treated with 40% TFA/$CH_2Cl_2$ solution for 2 h. The precipitate was filtered, washed with $CH_2Cl_2$ and dried under vacuum to give 16.5 mg (67%) of **5c**. $^1$H NMR (500 MHz, DMSO-$d_6$) δ 7.85 (m, aromatic, 2H), 7.69 (m, aromatic, 1H), 7.62 (d, C*H*$^{\beta}$=CH$^{\alpha}$, J=16.0 Hz, 1H), 7.51 (m, aromatic, 1H), 6.53 (d, CH$^{\beta}$=C*H*$^{\alpha}$, J=16.0 Hz, 1H).

**m-Sulfocinnamic acid (5d)**[44] and **3-amino-1,2-benzenediol (8e)**[45] were synthesized as previously described.

### *2.5 References*


1 Lunde, B. M.; Moore, C.; Varani, G. RNA-binding proteins: modular design for efficient function, *Nat Rev Mol Cell Bio* **2007**, *8*, 479.

2 Apponi, L. H.; Corbett, A. H.; Pavlath, G. K. RNA-binding proteins and gene regulation in myogenesis, *Trends Pharmacol Sci* **2011**, *32*, 652.

3 Keene, J. D. RNA regulons: coordination of post-transcriptional events, *Nat Rev Genet* **2007**, *8*, 533.

4 van Kouwenhove, M.; Kedde, M.; Agami, R. MicroRNA regulation by RNA-binding proteins and its implications for cancer, *Nat Rev Cancer* **2011**, *11*, 644.

5 Hermann, T. Strategies for the Design of Drugs Targeting RNA and RNA-Protein Complexes, *Angew Chem Int Ed Engl* **2000**, *39*, 1890.

6 Gallego, J.; Varani, G. Targeting RNA with small-molecule drugs: therapeutic promise and chemical challenges, *Accounts Chem Res* **2001**, *34*, 836.

7 Stockwell, B. R. Chemical genetics: ligand-based discovery of gene function, *Nat Rev Genet* **2000**, *1*, 116.

8 Schreiber, S. L. Small molecules: the missing link in the central dogma, *Nat Chem Biol* **2005**, *1*, 64.

9 Spring, D. R. Chemical genetics to chemical genomics: small molecules offer big insights, *Chem Soc Rev* **2005**, *34*, 472.

10 Thomas, J. R.; Hergenrother, P. J. Targeting RNA with small molecules, *Chem Rev* **2008**, *108*, 1171.

11 Cho, D. H.; Tapscott, S. J. Myotonic dystrophy: Emerging mechanisms for DM1 and DM2, *Biochim Biophys Acta - Molecular Basis of Disease* **2007**, *1772*, 195.

12 Brook, J. D.; McCurrach, M. E.; Harley, H. G.; Buckler, A. J.; Church, D.; Aburatani, H.; Hunter, K.; Stanton, V. P.; Thirion, J. P.; Hudson, T.; et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member, *Cell* **1992**, *69*, 385.

13 Gareiss, P. C.; Sobczak, K.; McNaughton, B. R.; Palde, P. B.; Thornton, C. A.; Miller, B. L. Dynamic combinatorial selection of molecules capable of inhibiting the (CUG) repeat RNA-MBNL1 interaction in vitro: discovery of lead compounds targeting myotonic dystrophy (DM1), *J Am Chem Soc* **2008**, *130*, 16254.

14 Arambula, J. F.; Ramisetty, S. R.; Baranger, A. M.; Zimmerman, S. C. A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding, *Proc Natl Acad Sci U S A* **2009**, *106*, 16068.

15 Pushechnikov, A.; Lee, M. M.; Childs-Disney, J. L.; Sobczak, K.; French, J. M.; Thornton, C. A.; Disney, M. D. Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3, *J Am Chem Soc* **2009**, *131*, 9767.

16 Warf, M. B.; Nakamori, M.; Matthys, C. M.; Thornton, C. A.; Berglund, J. A. Pentamidine reverses the splicing defects associated with myotonic dystrophy, *Proc Natl Acad Sci U S A* **2009**, *106*, 18551.

17 Hamy, F.; Gelus, N.; Zeller, M.; Lazdins, J. L.; Bailly, C.; Klimkait, T. Blocking HIV replication by targeting Tat protein, *Chem Biol* **2000**, *7*, 669.

18 Cheng, K.; Wang, X.; Yin, H. Small-Molecule Inhibitors of the TLR3/dsRNA Complex, *J Am Chem Soc* **2011**, *133*, 3764.

19 Gayle, A. Y.; Baranger, A. M. Inhibition of the U1A-RNA complex by an aminoacridine derivative, *Bioorg Med Chem Lett* **2002**, *12*, 2839.

20 Meisner, N. C.; Hintersteiner, M.; Mueller, K.; Bauer, R.; Seifert, J. M.; Naegeli, H. U.; Ottl, J.; Oberer, L.; Guenat, C.; Moss, S.; Harrer, N.; Woisetschlaeger, M.; Buehler, C.; Uhl, V.; Auer, M. Identification and mechanistic characterization of low-molecular-weight inhibitors for HuR, *Nat Chem Biol* **2007**, *3*, 508.

21 Chae, M. J.; Sung, H. Y.; Kim, E. H.; Lee, M.; Kwak, H.; Chae, C. H.; Kim, S.; Park, W. Y. Chemical inhibitors destabilize HuR binding to the AU-rich element of TNF-alpha mRNA, *Exp Mol Med* **2009**, *41*, 824.

22 Meisner, N. C.; Hintersteiner, M.; Seifert, J. M.; Bauer, R.; Benoit, R. M.; Widmer, A.; Schindler, T.; Uhl, V.; Lang, M.; Gstach, H.; Auer, M. Terminal adenosyl transferase activity of posttranscriptional regulator HuR revealed by confocal on-bead screening, *J Mol Biol* **2009**, *386*, 435.

23 Messias, A. C.; Sattler, M. Structural basis of single-stranded RNA recognition, *Accounts Chem Res* **2004**, *37*, 279.

24 Handa, N.; Nureki, O.; Kurimoto, K.; Kim, I.; Sakamoto, H.; Shimura, Y.; Muto, Y.; Yokoyama, S. Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein, *Nature* **1999**, *398*, 579.

25 Deshpande, G.; Samuels, M. E.; Schedl, P. D. Sex-lethal interacts with splicing factors in vitro and in vivo, *Mol Cell Biol* **1996**, *16*, 5036.

26 Kanaar, R.; Lee, A. L.; Rudner, D. Z.; Wemmer, D. E.; Rio, D. C. Interaction of the sex-lethal RNA binding domains with RNA, *EMBO J* **1995**, *14*, 4530.

27 Samuels, M.; Deshpande, G.; Schedl, P. Activities of the Sex-lethal protein in RNA binding and protein:protein interactions, *Nucleic Acids Res* **1998**, *26*, 2625.

28 Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C. H.; Nagai, K. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature* **1994**, *372*, 432.

29 Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem* **2004**, *25*, 1605.

30 Seidler, J.; McGovern, S. L.; Doman, T. N.; Shoichet, B. K. Identification and prediction of promiscuous aggregating inhibitors among known drugs, *J Med Chem* **2003**, *46*, 4477.

31 Feng, B. Y.; Shoichet, B. K. A detergent-based assay for the detection of promiscuous inhibitors, *Nat Protoc* **2006**, *1*, 550.

32 Owicki, J. C. Fluorescence polarization and anisotropy in high throughput screening: perspectives and primer, *J Biomol Screen* **2000**, *5*, 297.

33 Hellman, L. M.; Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions, *Nature protocols* **2007**, *2*, 1849.

34 Coan, K. E.; Shoichet, B. K. Stability and equilibria of promiscuous aggregates in high protein milieus, *Mol Biosyst* **2007**, *3*, 208.

35 Rishi, V.; Potter, T.; Laudeman, J.; Reinhart, R.; Silvers, T.; Selby, M.; Stevenson, T.; Krosky, P.; Stephen, A. G.; Acharya, A.; Moll, J.; Oh, W. J.; Scudiero, D.; Shoemaker, R. H.; Vinson, C. A high-throughput fluorescence-anisotropy screen that identifies small molecule inhibitors of the DNA binding of B-ZIP transcription factors, *Anal Biochem* **2005**, *340*, 259.

36 Bond, A.; Reichert, Z.; Stivers, J. T. Novel and specific inhibitors of a poxvirus type I topoisomerase, *Mol Pharmacol* **2006**, *69*, 547.

37 Seiple, L. A.; Cardellina, J. H., 2nd; Akee, R.; Stivers, J. T. Potent inhibition of human apurinic/apyrimidinic endonuclease 1 by arylstibonic acids, *Mol Pharmacol* **2008**, *73*, 669.

38 Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays, *J Med Chem* **2010**, *53*, 2719.

39 Poilane-Dykas, L. H., Ph.D., Dissertation, Wesleyan University, 2008.

40 Heyerdahl, S. L.; Rozenberg, J.; Jamtgaard, L.; Rishi, V.; Varticovski, L.; Akah, K.; Scudiero, D.; Shoemaker, R. H.; Karpova, T. S.; Day, R. N.; McNally, J. G.; Vinson, C. The arylstibonic acid compound NSC13746 disrupts B-ZIP binding to DNA in living cells, *Eur J Cell Biol* **2010**, *89*, 564.

41 Kim, H.; Cardellina, J. H., 2nd; Akee, R.; Champoux, J. J.; Stivers, J. T. Arylstibonic acids: novel inhibitors and activators of human topoisomerase IB, *Bioorg Chem* **2008**, *36*, 190.

42 El-Batta, A.; Jiang, C.; Zhao, W.; Anness, R.; Cooksy, A. L.; Bergdahl, M. Wittig reactions in water media employing stabilized ylides with aldehydes. Synthesis of alpha,beta-unsaturated esters from mixing aldehydes, alpha-bromoesters, and Ph3P in aqueous NaHCO3, *J Org Chem* **2007**, *72*, 5244.

43 Muthukumaran, K.; Loewe, R. S.; Ambroise, A.; Tamaru, S.; Li, Q.; Mathur, G.; Bocian, D. F.; Misra, V.; Lindsey, J. S. Porphyrins bearing arylphosphonic acid tethers for attachment to oxide surfaces, *J Org Chem* **2004**, *69*, 1444.

44 Muller, C. E.; Sandoval-Ramirez, J.; Schobert, U.; Geis, U.; Frobenius, W.; Klotz, K. N. 8-(Sulfostyryl)xanthines: water-soluble A2A-selective adenosine receptor antagonists, *Bioorg Med Chem* **1998**, *6*, 707.

45 Corfield, J. A.; Grimes, R. M.; Harrison, D.; Hartley, C. D.; Howes, P. D.; Le, J.; Meeson, M. L.; Mordaunt, J. E.; Shah, P.; Slater, M. J.; White, G. V. 2007.

# CHAPTER 3

# Electrostatic Interactions and Conformational Selection in Two-step U1A-SL2 RNA Dissociation Kinetics

The work in this chapter would not have been possible without the help of Irisbel Guzman, who has contributed greatly to this project.

## *3.1 Introduction*

### *3.1.1 Dynamics of Protein-RNA complex formation*

In nearly all protein-RNA complexes that have been characterized, structural changes are seen in both protein and RNA. Ribosomal proteins, including L11 and L30 have been shown to exhibit an induced fit mode of binding. The target RNAs of these proteins, also show a high degree of conformational change upon binding. NMR studies and fluorescent studies using Trp mutants of human tristetrapolin have suggested that both induced fit and conformational selection are important for the protein to bind the target AU rich RNA.[1,2] In the case of aminoacyl-tRNA synthetase, precise recognition of the cognate tRNA is achieved by electrostatic interactions with discriminating functional groups on the tRNA bases as well as the dynamics of the protein and tRNA.[3] Sequence dependent tRNA conformations are specifically recognized by the protein, and further conformational adjustments of the protein have been observed.[3] Together, these examples emphasize that the dynamics of the protein and the RNA are crucial factors for determining the binding mechanism, affinity, and specificity of the protein-RNA complexes.

*3.1.2 Dynamics of U1A-SL2 RNA complex*

Helix C, loop1 and loop 3 of the U1A protein undergo conformational rearrangements to form the complex. The NMR structure of the free U1A protein shows that helix C is positioned across the β-sheet, which is the primary RNA binding surface. In the U1A-RNA complex, helix C is moved away from the β-sheet surface, allowing the residues in the β-sheet to contact the SL2 RNA. Several experiments have suggested that although the NMR structure has captured the closed helix C conformation, in solution the helix C may be dynamic.[4,5] Jean and coworkers have suggested that the movement of helix C is on a nanosecond or a longer timescale.[5]

Loop 3 is flexible in the free form and becomes rigid upon binding to the RNA, and protudes through the loop region of the SL2.[6,7] This conformational change presents the RNA bases to the β-surface of the protein forming a stable complex. Molecular dynamics (MD) simulations have shown that the mutation of Gly53, which is located at the junction of loop 3 and β3, to a Val significantly reduces the mobility of loop 3 and results in a lower binding affinity.[8] In addition, correlated motions of loop1, loop 3 and helix C are proposed to be important for the U1A-SL2 complex formation.[8]

Although the structure of the free RNA has not been determined, NMR studies have suggested that the chemical shifts of many protons for bound RNA are substantially different from those for free RNA, especially in the loop region of the RNA. Molecular dynamics simulations have shown the SL2 RNA is dynamic in the absence of the U1A protein and becomes more ordered upon binding to the RNA.[9-11] The bases of the loop region are mostly, but not always, facing the interior of the RNA. Therefore, a conformational selection may be occurring to some extent during the association with the protein.

*3.1.3 Kinetics of U1A-SL2 RNA complex dissociation*

Dynamic processes involved in the association and dissociation of protein-RNA complexes can be probed by detailed kinetic studies. The kinetics of association of the U1A and SL2 RNA complex has been studied using surface plasmon resonance (SPR).[12-15] By comparing the rate of association and dissociation of the complexes of the wild type and mutant proteins, a 'lure and lock' mechanism of binding was proposed.[12] Like other macromulecular recognition processes, an initial association of the complex that is dominated by electrostatic interactions is followed by conformational rearrangements to form the tight complex. Mutation of positively charged residues increased the rate of association and the mutation of residues that have key stacking interaction with the RNA decreased the rate of dissociation. This data supports the importance of positively charged residues in the initial 'lure' step and the stacking residues in complex stability. Additional SPR studies have each revealed the role of the Tyr13-Phe56-Tyr13 network and loop 3 in the U1A-SL2 association pathway.

More recently, Anunciado *et al.* have developed a two-step pathway for the dissociation between U1A and SL2 RNA based on data obtained from temperature jump and stopped flow experiments.[16] In this pathway the first fast step of ~100 μs is the loosening of the U1A-SL2 RNA complex and the second slow step of ~233 ms is the actual dissociation of the complex.

*3.1.4 Electrostatic interactions in protein-RNA complexes*

Electrostatic interactions mediated by basic residues in RNA binding proteins are important for overall protein-RNA binding affinity and may also play a role in specificity of these complexes.[17] Indeed, a computational study has shown that the largest overlap between a positive electrostatic patch and the binding surface is observed with a RNA binding protein, L1

ribosomal protein.[18] Large and small positively charged surfaces of RNA binding proteins are critical during short- and long-range interactions that lead to binding. A positive patch in the center of the aminoacyl-tRNA synthetase has been suggested to play an important role in long-range interactions with the cognate tRNA, which is the decisive factor of primary association.[19,20] The aminoacyl-tRNA synthetase and the tRNA are brought together when the negative potential of the tRNA fits the positive potential patch of the enzyme.[20]

The net charge of the N-terminal RRM of the U1A protein is +7 while that of the C-terminal RRM is 0.[21] Many of the positively charged residues are missing in the C-terminal RRM or replaced by a negatively charged or a neutral residue. This difference may contribute to the fact that the C-terminal RRM does not bind to RNA, while the N-terminal RRM binds to SL2 RNA with a sub-nanomolar $K_d$.[22] A similar trend was found in a mouse neural protein, Musashi1. Musashi1 contains two RRMs, but only the N-terminal RRM binds to the target RNA with high specificity.[23] The positive potential is prevailing for the β-sheet surface of RRM1, while the β-sheet surface of RRM2 is largely neutral with small negative potential patches. Considering the fact that the overall structural fold is essentially identical for RRM1 and RRM2, the positive electrostatic potential of the RRM1 is thought to play a significant role in RNA binding.[23]

The role of basic residues in the N peptide (derived from N protein in phage λ) and box B RNA complex has been studied using parallel experimental measurements and theoretical calculations. Calculations based on the non-linear Poisson-Boltzmann equation revealed that all basic residues of the peptide contribute significantly to the overall binding free energy, but the relative contributions vary depending on the environment of a residue.[17]

Here, we studied the role of four lysines during the two-step U1A-SL2 RNA complex dissociation kinetics. We performed laser-induced temperature jump and stopped-flow kinetics experiments with tryptophan-labeled mutants of U1A. We found that Lys20, Lys23 and Lys50 are crucial for the formation of a tight complex between U1A and SL2 RNA: the fast phase ($\tau_1$) associated with tight binding disappears when these residues were mutated to glutamine. Stopped flow kinetics of these mutants reveals no significant change in the slow phase ($\tau_2$) associated with dissociation of the loosely bound state. In contrast, mutating Lys22 had no effect on the fast kinetics of the tight complex, but the slow phase ($\tau_2$) sped up, implying that Lys22 is important for the stability of the loose complex. Interestingly, a fast phase of ~27 ms with negative amplitude was resolved in a temperature jump experiment with U1A Lys22Gln without RNA. We propose that this negative fast phase monitors a conformational transition of U1A similar to that which occurs upon tight RNA binding. If so, the Lys22Gln mutant is closer to conformational selection than the wild type in the continuum from conformational selection to induced fit.

## *3.2 Results*

### *3.2.1 Selection of mutated positively charged residues*

We used a 101-amino acid N-terminal U1A fragment that has been shown to be sufficient for high affinity binding to SL2 RNA. The Trp fluorescent label for the kinetics studies was introduced into the Phe56 position. Phe56 forms a stacking interaction with A6 of the SL2 RNA in the complex.[7] This Phe56Trp mutation was previously reported to have only minimal effects on binding affinity and was successfully used to monitor the dissociation of the U1A-SL2

complex.[5,24] The titration of SL2 RNA to the U1A (F56W) solution quenches the Trp fluorescence presumably due to the stacking interaction between the Trp and the A6.[16]

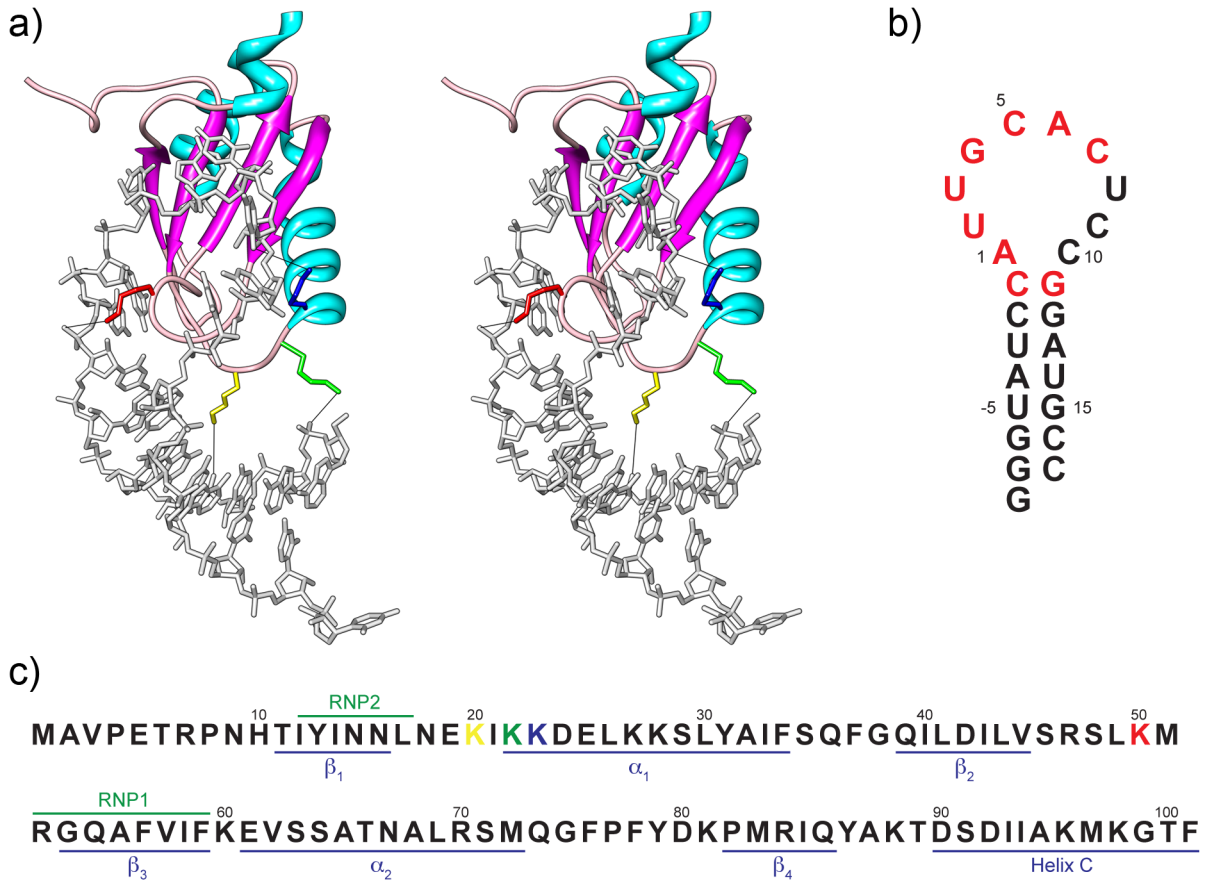



**Figure 3.1** a) Stereoview structure of the U1A-SL2 complex. Mutant residues are indicated in different colors. K20 is shown in yellow, K22 in green, K23 in blue and K50 in red. Black solid lines show the electrostatic interactions of the lysine residues with the closest phosphate group in the SL2 RNA. Distances from each lysine residue to the closest phosphate group of SL2 RNA are K20: 6.6Å (C-2); K22: 4.1Å (A-4); K23: 7.7Å (C9); K50: 7.7Å (U3). The structure was generated with the program Chimera.[25] b) Sequence of SL2 RNA. Nucleotides that are directly recognized by the protein are shown in red. c) Sequence of the N-terminal RRM of U1A protein.

To analyze the role of electrostatic interactions in the two-step dissociation pathway observed with U1A (Phe56Trp)-SL2 complex, four lysine residues, Lys20, Lys22, Lys23 and Lys50, presented in Figure 3.1a and c, that have electrostatic interaction with RNA, but do not

form hydrogen bond with the RNA, were chosen for mutation. The residues were mutated to glutamine to conserve the relative size of the residue while eliminating the positive charge.

Previous studies have shown that all four of these residues have an effect on the association of the complex using surface plasmon resonance.[12,15] Whereas other positively charged residues that are located on the top of the protein (Arg7, Lys60 and Arg70) did not show a significant difference from the wild type.[15] In the U1A-SL2 complex, Lys20 is close to the upper part of the RNA stem, while Lys 22 lies near the bottom tip of the RNA stem. Lys23 has electrostatic interaction with the phosphate groups of C9 and C10 in the loop region. Lys50 also has electrostatic interactions with the phosphate groups in the loop region. However it is closest to the phosphate groups of U3 and G4. Therefore, these four residues participate in electrostatic interactions with different parts of the RNA phosphate backbone.

*3.2.2 Determination of optimal dissociation conditions*

To determine the optimal conditions for the kinetic analysis with stopped-flow and temperature-jump, we measured the equilibrium binding affinity of each of the mutant U1A proteins using an electrophoretic mobility shift assay (Table 1). At 25 °C and 200 mM KCl, all of the Lys to Gln mutations decreased the binding affinity compared to the F56W mutant. The mutation of Lys22 and Lys50 each decreased the binding affinity more than 14-fold. Lys20Gln and Lys23Gln had a less significant effect on the binding affinity, only a ~3-fold and ~6-fold decrease, respectively. The same trend was previously observed by another lab.[7]

For the temperature jump experiment, we analyzed the binding affinity at 25 °C and 35 °C. The higher temperature decreased the binding affinity of U1A (Phe56Trp) ~5-fold, which is

sufficient to induce dissociation. The other mutants also showed decreased binding affinity at higher temperature.

In the stopped-flow setup, two solutions are rapidly mixed together. For complex dissociation, one syringe was filled with the protein-RNA complex, and the other was filled with buffer. Therefore either the binding affinity of the complex has to be weak enough to dissociate with two-fold dilution, or we can use a high concentration of salt in the buffer syringe to drive dissociation. Based on the binding affinity of U1A (Phe56Trp), we decided to jump the salt concentration from 0.2 to 1 M KCl for dissociation. The increase of salt concentration decreased the binding affinity from 22 (±1) nM to 110 (±20) nM, sufficient for dissociation. The effect of KCl concentration on the $K_d$ values of the other mutants was smaller. The loss of binding affinity ranged from 1.3-fold to 3.3-fold. The two-fold dilution of the complex solution and the slight decrease in binding affinity at 1 M KCl still dissociated those complexes, as shown in the stopped-flow section.

**Table 3.1** $K_d$ values of mutant U1A proteins.

| Protein | 200 mM KCl 25°C | 200 mM KCl 35°C | 1 M KCl 25°C |
|---|---|---|---|
| F56W | 2.21 (±0.09) × $10^{-8}$ | 1.2 (±0.3) × $10^{-7}$ | 1.1 (±0.2) × $10^{-7}$ |
| K20Q/F56W | 7.4 (±0.5) × $10^{-8}$ | 5 (±1) × $10^{-7}$ | 2.4 (±0.2) × $10^{-7}$ |
| K22Q/F56W | 3.28 (±0.06) × $10^{-7}$ | 1.2 (±0.2) × $10^{-6}$ | 4.3 (±0.2) × $10^{-7}$ |
| K23Q/F56W | 1.32 (±0.04) × $10^{-7}$ | 3.7 (±0.2) × $10^{-7}$ | 3.53 (±0.03) × $10^{-7}$ |
| K50Q/F56W | 3.92 (±0.05) × $10^{-7}$ | 2.2 (±0.5) × $10^{-6}$ | 5.8 (±0.8) × $10^{-7}$ |

*3.2.3 Circular dichroism measurements*

To confirm that the secondary structure does not change upon temperature jump we evaluated the structures of the U1A mutants using circular dichroism. Each mutant was measured at 25 °C and 45 °C, because the higher temperature jump was performed from 35 °C to 45 °C. The CD spectra for all of the mutants were similar to the pseudo wild type control (Phe56Trp) at each temperature and overlapped almost perfectly between the two temperatures (Figure 3.2), indicating that the substitution of Lys for Gln and the temperature jump do not introduce perturbation of the secondary structure of the protein. Therefore, any fast phase observed during temperature jump is not due to protein unfolding.

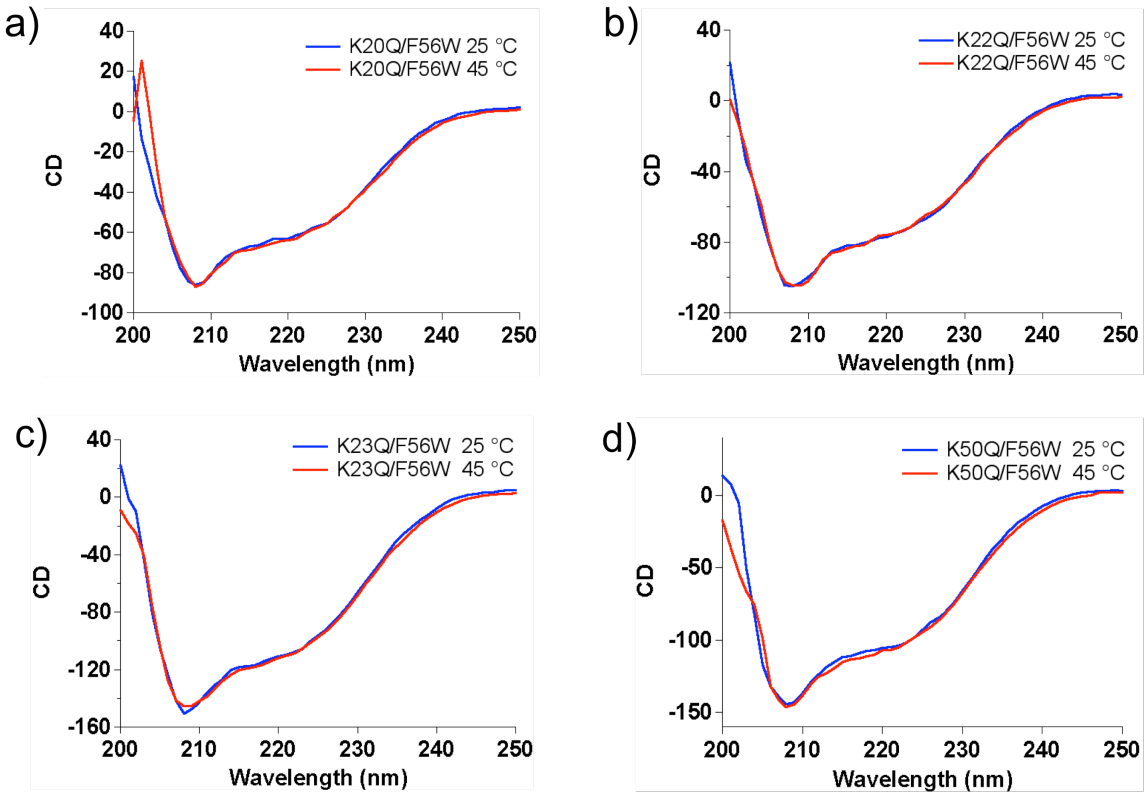



**Figure 3.2** Circular dichroism of U1A mutants. a) K20Q/F56W b) K22Q/F56W c) K23Q/F56W d) K50Q/F56W. 10 μM solution of each proteins was measured in sodium cacodylate buffer.

*3.2.4 Temperature jump measurements*

To monitor the earliest step of RNA-protein complex dissociation, Irisbel Guzman performed laser-induced temperature jump experiments. Our probe signal was the Trp fluorescence decay change $\chi$, normalized from 1 (pre-jump) to 0 (post-jump). The protein-RNA complex solution in a 200 mM KCl buffer was perturbed by temperature jumps from 25 °C to 35 °C and from 35 °C to 45 °C (See Methods for more details of data analysis). A single phase with a relaxation time of ~35 μs and a positive amplitude was observed for the Phe56Trp mutant complex (Figure 3.3).

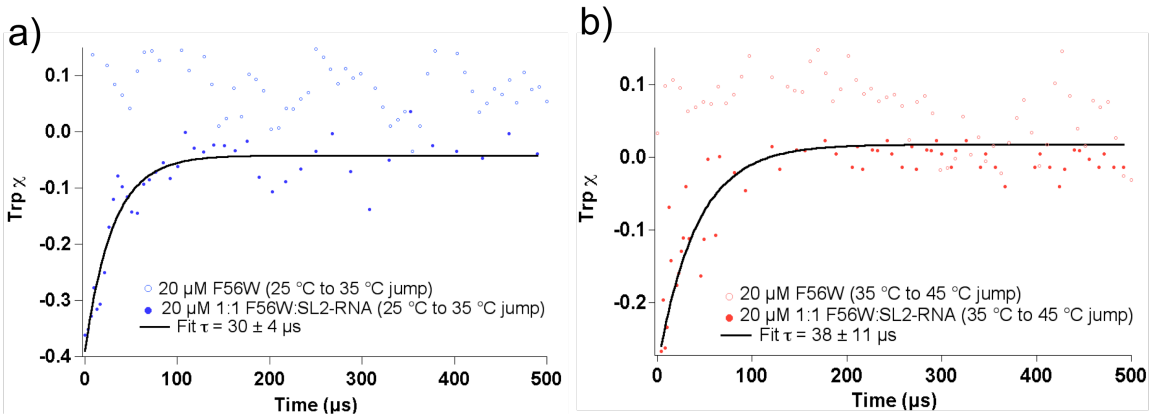



**Figure 3.3** Temperature jump of 20 μM F56W a) From 25 °C to 35 °C b) From 35 °C to 45 °C

The Lys22Gln/Phe56Trp mutation had no effect on the fast phase, other than a slightly longer relaxation time (Figure 3.4c and d). The fast phase did not significantly change with temperature between 35 and 45 °C. This is consistent with the same transition from a tight complex to a loosely bound RNA-protein complex that was observed for Phe56Trp here and previously reported by Anunciado *et al.*[16] Our observation implies that the positive charge of Lys22 is not required for tight complex formation.

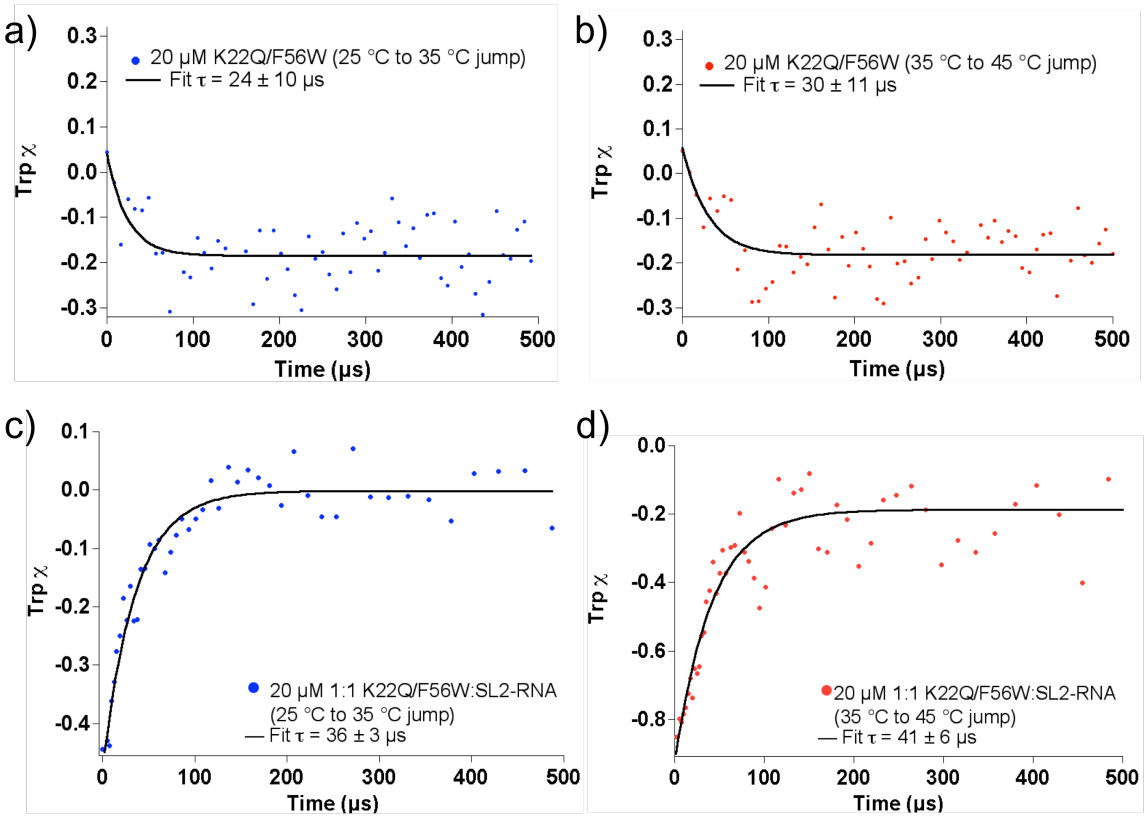



**Figure 3.4** Temperature jumps of K22Q/F56W mutant in 200 µM KCl buffer. a) T-jump of just K22Q/F56W from 25 °C to 35 °C. b) T-jump of just K22Q/F56W from 35 °C to 45 °C for just 20 µM F56W mutant. c) T-jump of 20 µM K22Q/F56W-SL2 RNA complex from 25 °C to 35 °C. d) T-jump of 20 µM K22Q/F56W-SL2-RNA from 35 °C to 45 °C.

However, we did not observe the fast phase with the Lys20Gln/Phe56Trp, Lys23Gln/Phe56Trp and Lys50Gln/Phe56Trp mutants (Figure 3.5). Therefore these mutations disrupt the equilibrium between the tightly and loosely bound complex. As control experiments, Lys20Gln/Phe56Trp, Lys23Gln/Phe56Trp and Lys50Gln/Phe56Trp mutants were temperature jumped without RNA, and no fast phase was observed. In contrast, the temperature jump of Lys22Gln/Phe56Trp with no RNA has a negative fast phase of ~27 µs shown in Figure 3.4a and b. Although this phase has the opposite amplitude, it has a similar time scale as the conversion between tight and loose U1A-SL2 complexes. The fast negative phase in Lys22Gln/Phe56Trp is

evidence for a conformation transition of the protein on a similar time scale as the conformational transformation upon loose-tight bound complex interconversion.

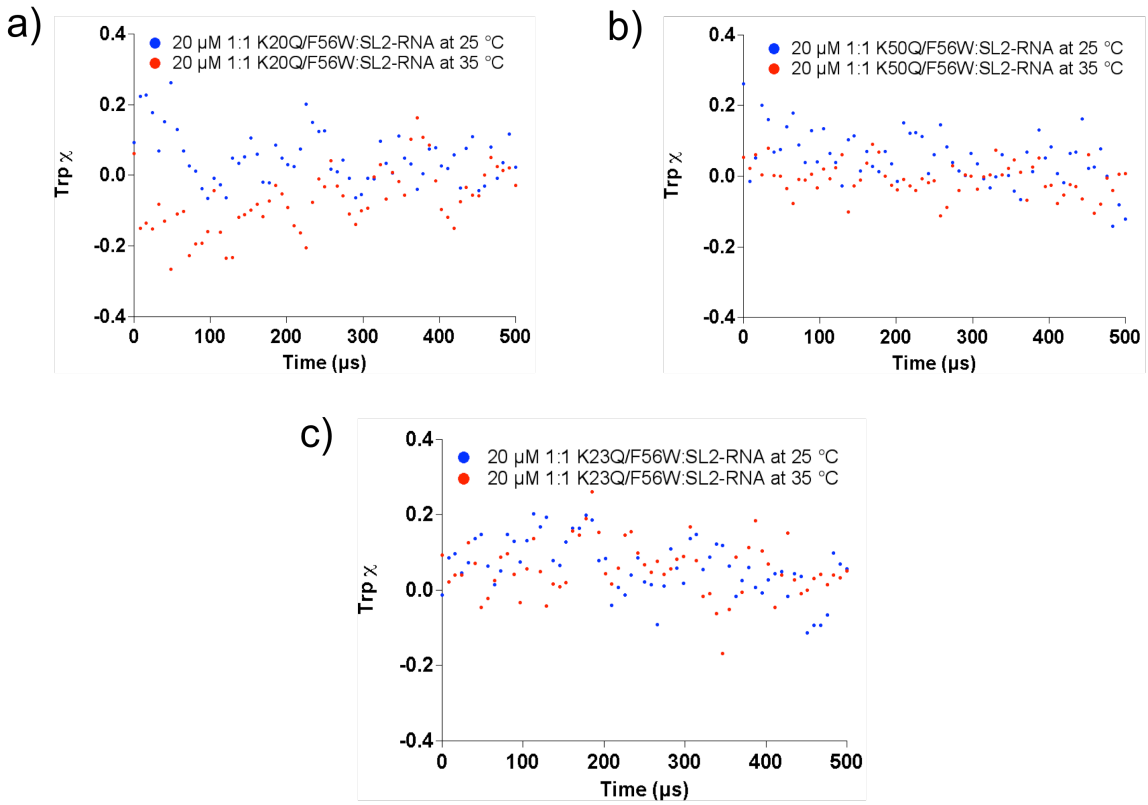



**Figure 3.5** Temperature jump of 20 μM U1A-SL2 RNA complexes at 25 °C and 35 °C. a) K20Q/F56W b) K23Q/F56W c) K50Q/F56W

*3.2.5 Stopped-flow kinetics measurements*

To measure the complete dissociation kinetics for the positively charged mutants, Irisbel Guzman performed stopped-flow measurements of protein-RNA complexes. Try56 fluorescence intensity between 320-450 nm was monitored as the probe. The stopped-flow results after fast mixing of RNA-protein complex with 1 M KCl buffer are shown in Figures 3.6 to Figure 3.9. Under these conditions, the dissociation rate should dominate the relaxation process. A single phase on the millisecond time scale was observed upon fast mixing for the complex dissociation of Lys20Gln/Phe56Trp, Lys22Gln/Phe56Trp, Lys23Gln/Phe56Trp and Lys50Gln/Phe56Trp

mutants. The dissociation rate of the Lys22Gln/Phe56Trp-SL2 complex was somewhat faster than that of the other mutants and the pseudo wild type control (Table 3.2).

The fluorescence amplitude increased almost to the protein-only baseline for all of the mutants, indicating nearly full dissociation of the complex. The data was fit by a single exponential function and $\tau_2$ was not concentration-dependent (Table 3.2). These observations are consistent with the unimolecular complex dissociation rate dominating over bimolecular association kinetics under our buffer conditions. As shown in Figure 3.7 and Figure 3.9, the dissociation time $\tau_2$ increases 10-fold upon heating by 10 °C for all the mutants.

**Table 3.2** Relaxation time ($\tau_2$) of U1A mutants measured by sttoped-flow.

| Protein | 1 μM | | 0.5 μM | |
|---|---|---|---|---|
| | 25°C | 35°C | 25°C | 35°C |
| F56W | 2.03 (±0.01) | 0.23 (±0.02) | 2.31 (±0.02) | 0.26 (±0.01) |
| K20Q/F56W | 1.29 (±0.01) | 0.12 (±0.02) | 1.29 (±0.02) | 0.13 (±0.04) |
| K22Q/F56W | 0.89 (±0.03) | 0.086 (±0.002) | 0.96 (±0.03) | 0.088 (±0.002) |
| K23Q/F56W | 1.80 (±0.03) | 0.19 (±0.04) | 2.02 (±0.05) | 0.19 (±0.05) |
| K50Q/F56W | 1.16 (±0.02) | 0.11 (±0.03) | 1.19 (±0.02) | 0.10 (±0.02) |

*Unit is second. $\tau_2$ was determined using single exponential equation.

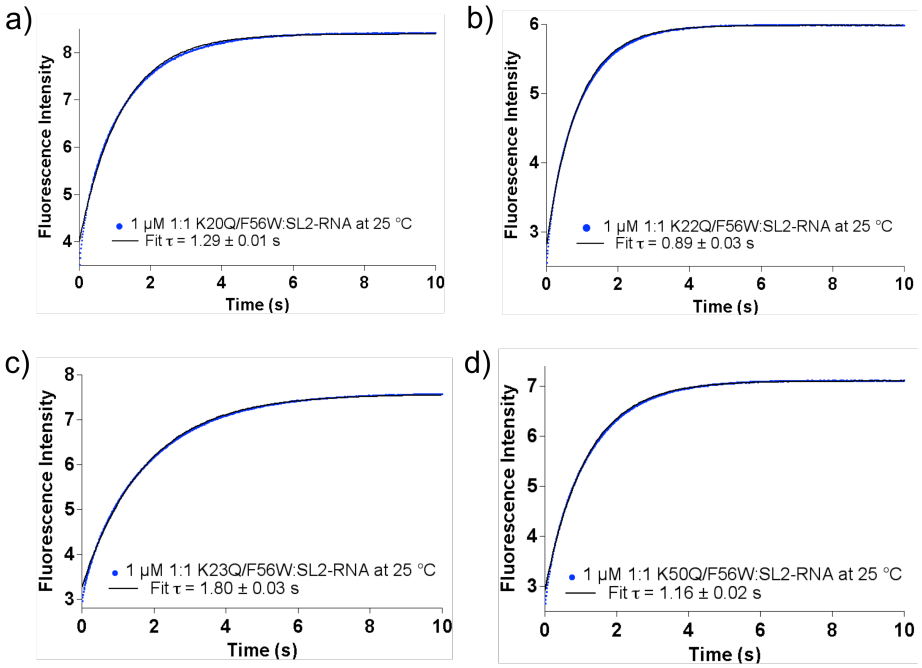



**Figure 3.6** Stopped-flow measurements of 1 μM U1A-SL2 RNA complexes at 25 °C. a) K20Q/F56W b) K22Q/F56W c) K23Q/F56W d) K50Q/F56W

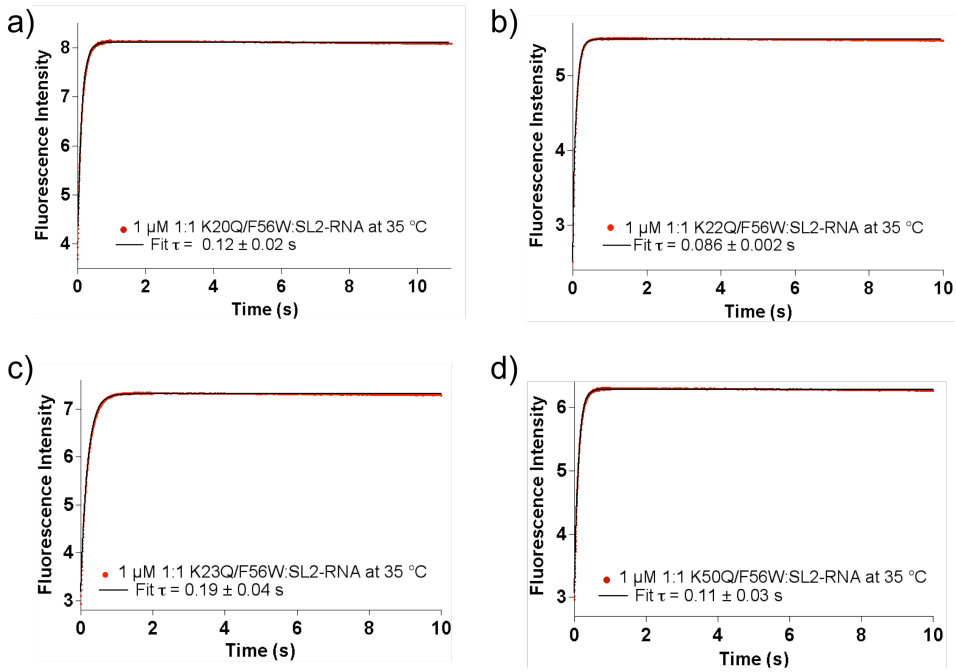



**Figure 3.7** Stopped-flow measurements of 1 μM U1A-SL2 RNA complexes at 35 °C. a) K20Q/F56W b) K22Q/F56W c) K23Q/F56W d) K50Q/F56W

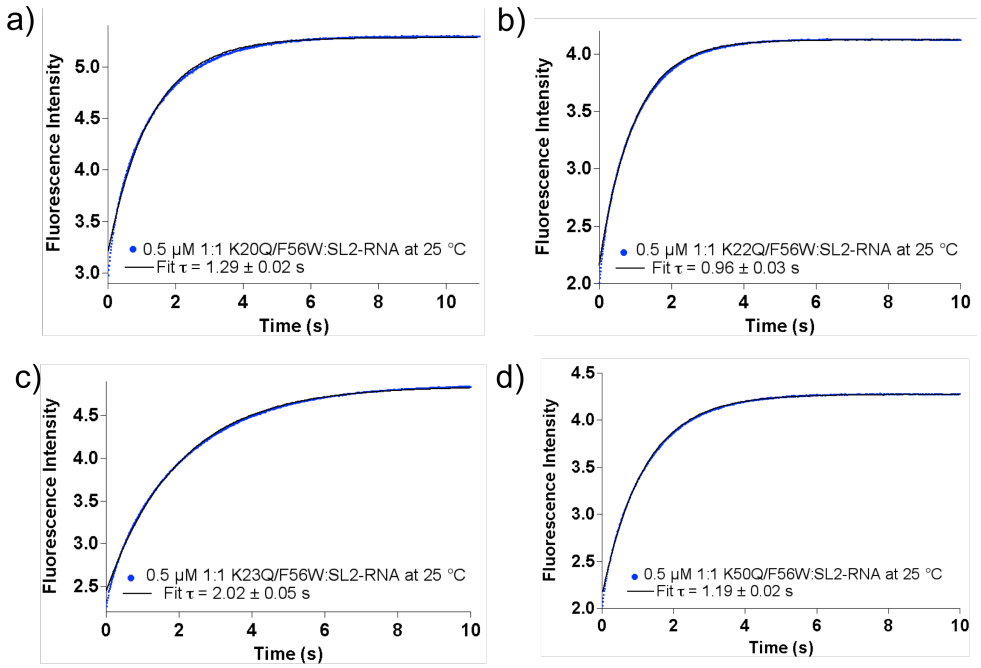



**Figure 3.8** Stopped-flow measurements of 0.5 μM U1A-SL2 RNA complexes at 25 °C.
a) K20Q/F56W b) K22Q/F56W c) K23Q/F56W d) K50Q/F56W.

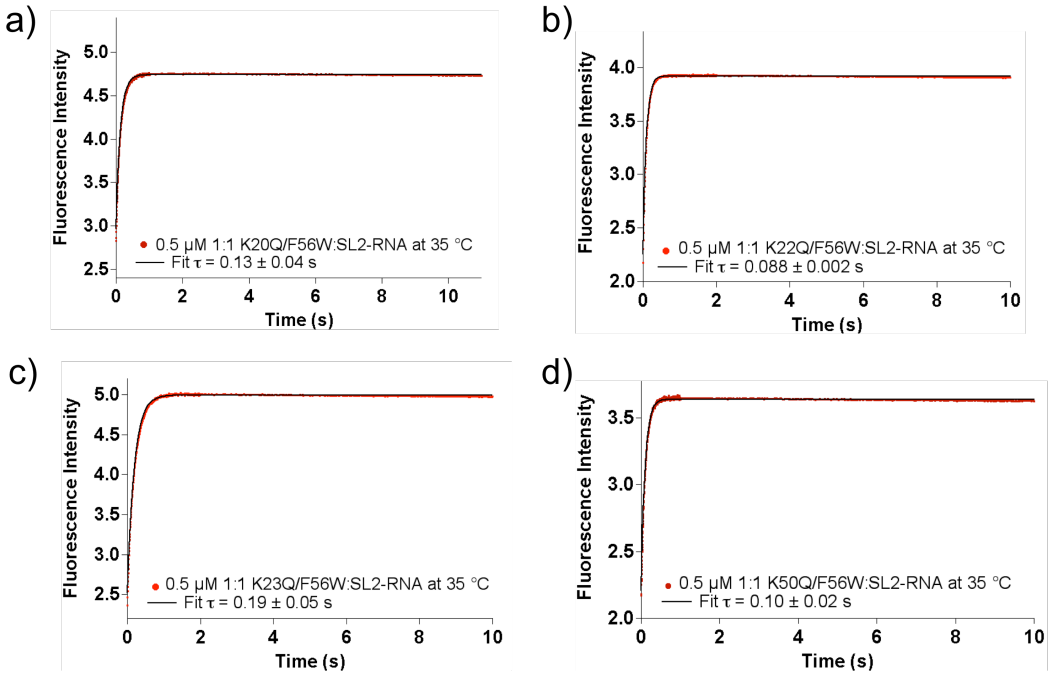



**Figure 3.9** Stopped-flow measurements of 0.5 μM U1A-SL2 RNA complexes in at 35 °C.
a) K20Q/F56W b) K22Q/F56W c) K23Q/F56W d) K50Q/F56W.

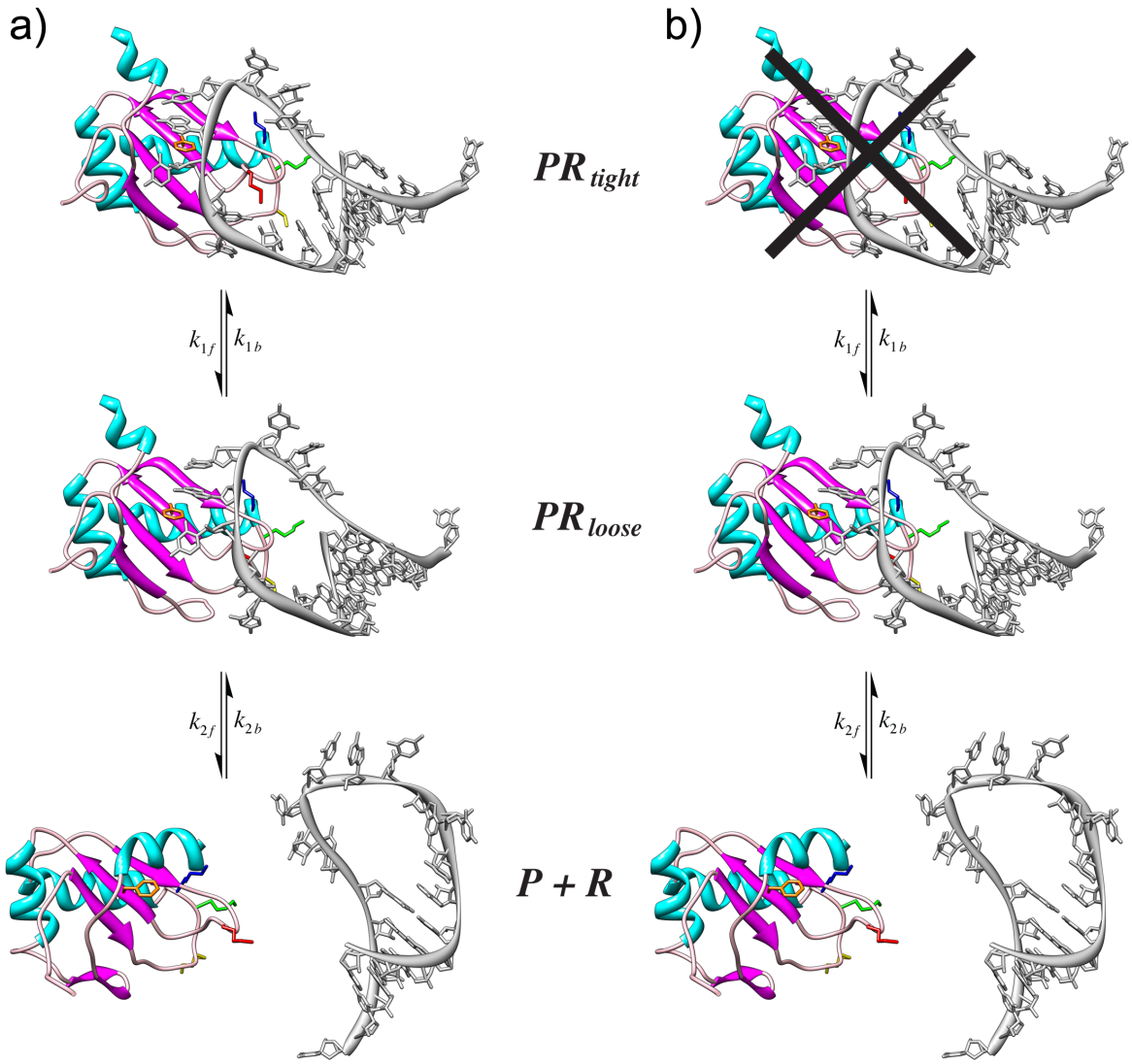



**Figure 3.10** Kinetic scheme for dissociation of the U1A-SL2 complex. a) K22Q mutation does not affect the overall kinetic scheme. b) K20Q, K23Q and K50Q mutants do not form the tight complex. Trp56 is shown in orange and the other mutations are indicated as in Figure 3.1.

*3.2.6 Global fit of the data*

To support our hypothesis of a two-step dissociation pathway ($PR_{tight} \rightarrow PR_{loose} \rightarrow P + R$), Professor Martin Gruebele performed a global fit of the kinetics and thermodynamics data for the F56W mutant. He used a similar approach as previously reported by Anunciado *et al.*[16] He simultaneously fitted the temperature jump, stopped-flow and EMSA binding data at all

temperatures measured. The global fit yielded activation free energies and the equilibrium free energies for each reaction step during dissociation, as well as linear free energy dependences on temperature and salt concentration, modeled by

$$\Delta G = \Delta G_0 + \Delta G_T (T - T_0) + \Delta G_{KCl}[KCl] \qquad \textbf{Eq. 1}$$

For the first step of dissociation ( $PR_{tight} \rightarrow PR_{loose}$ ) we obtained $\Delta G_0$ = 5.29 kJ/mole, and $\Delta G_T$ = -0.608 kJ/mole. The calculated $\Delta G_T$ shows that the tightly bound complex is destabilized by temperature faster than the loosely bound complex. For the second step ( $PR_{loose} \rightarrow P + R$ ), we obtained $\Delta G_0$ = 9.97 kJ/mole, $\Delta G_T$ = -0.07 kJ/mole, and $\Delta G_{KCl}$ = -23.1 kJ/mole. Thus complete dissociation is favored at higher temperatures and higher salt concentrations. The activation energies were fitted to equation

$$k_{AB} = k_0 e^{-\Delta G^{\dagger}_{AB}/RT} \qquad \textbf{Eq. 2}$$

The prefactor was fixed to $k_m$=(0.01 $\eta(T)$)$^{-1}$ ms$^{-1}$, using the normal Kramers viscosity dependence for both reactions.[26] The activation energies are $\Delta G^{\dagger}_{tight \rightarrow loose}$ = 8.76 kJ/mole and $\Delta G^{\dagger}_{loose \rightarrow dissociated}$ = 34.2 kJ/mole for the second step. No temperature dependence of the activation free energy was fitted because only two temperatures were measured. The actual barriers may be correspondingly higher if the diffusional prefactor is faster; if both reactions have a similar prefactor (as assumed here), complete dissociation has a significantly higher activation barrier than tight-loose interchange. The free energy landscape for tight binding and dissociation is shown in Figure 3.11.

### *3.3 Discussion*

Combining the temperature jump and stopped-flow data from pseudo wild type Phe56Trp and the Lys to Gln mutants we confirmed the two-step dissociation pathway previously proposed by our lab (Figure 3.10a). However for Lys20Gln, Lys23Gln, and Lys50Gln mutants a one-step dissociation pathway of loose complex to dissociated state was observed (3.10b). The disappearance of the fast transition between tight complex and loose complex observed with the glutamine mutants of Lys20, Lys23 and Lys50 demonstrate that the electrostatic interactions mediated by these residues are required for the tight complex formation between U1A and SL2 RNA. These residues are close to the loop region of the RNA where majority of the contacts are formed between the protein and RNA. These results suggest that the residues that contact the loop or close to the loop of the SL2 RNA are important for locking the RNA in position to form a tight complex.

For the Lys22Gln mutant, we observed a similar phase as the wild type control, which suggests that the relative energy of tight and loose energy states is similar to the wild type control (Figure 3.11a). Thus, the Lys22Gln mutation did not affect the formation of the tight complex. Lys22 may have a strong electrostatic interaction with A-5 of the RNA. The distance between Lys22 and the phosphate backbone of A-5 is approximately 5Å, while other lysine residues that were mutated are 7~9Å apart. Considering the fact that Lys22 may have the strongest electrostatic interaction with the protein, it is puzzling why neutralization of Lys22 did not affect the tight complex. The location of the residue may be the key to explain this result. Lys22 is contacting the furthest end of the RNA stem, which is not where the majority of the contacts are formed. Therefore we purpose that although this residue is important for the overall binding free energy, it is not crucial for the tight complex formation due to the region of the

RNA it interacts with. Previous SPR data has shown that Lys22 is crucial for complex stability.[15] Our stopped-flow results suggest that Lys22 contributes to the stability of the loose complex, holding the RNA and decreasing the rate of dissociation.

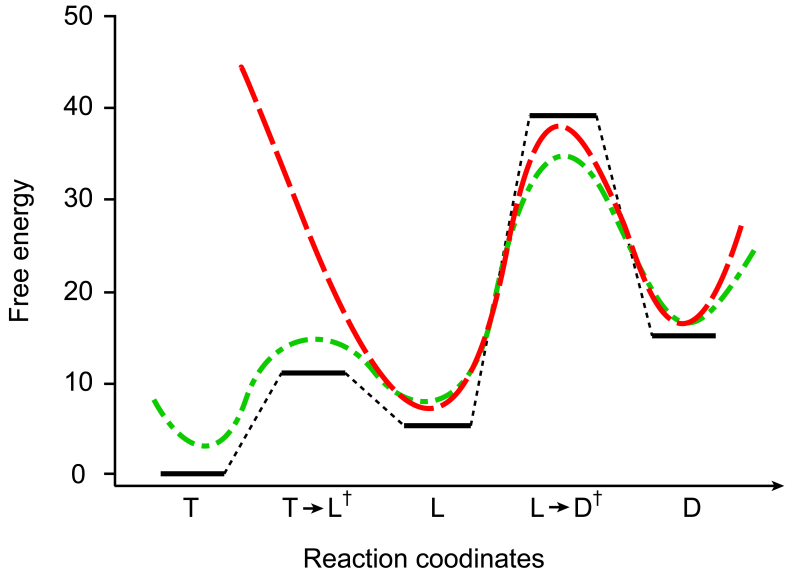



**Figure 3.11** Energy diagram of the two-step U1A-SL2 RNA complex dissociation. F56W is in black solid line, K22Q/F56W is in green dotted line, and the other mutants are in red dashed line. T is tight complex, L is loose complex and D is dissociated.

Binding of a protein to its substrate is often discussed in terms of conformational selection *vs*. induced fit in an either-or scenario.[27,28] Either the conformation "B" that the protein will occupy in the bound state is already low in free energy before the substrate binds; in that case the substrate can 'select' conformation "B", or the free energy of "B" is high relative to the conformation "A" that is occupied by the free protein before substrate binding; in that case the substrate 'induces' conformation "B." In reality, intermediate scenarios are possible as shown in Figure 3.12a.[27,28] In the simplest case where the conformational change is limited to the protein, we can describe the possible scenarios by a single parameter PSI (for Protein Selected or Induced):

$$\Psi = \frac{G_B(unbound) - G_A(unbound)}{RT} = \Delta G_{BA} \qquad \textbf{Eq. 3}$$

It should be clear that for a given substrate, Ψ is a property of the protein only: knowledge of the free energies of protein conformations A and B is required only in the unbound state. Numerically, Ψ describes the possible scenarios as follows: (1) Lock-and key: if $\Psi << 0$, state "B" is already lower in free energy and the substrate simply binds to the protein without inducing a conformational change; state "A" is then irrelevant. (2) Conformational selection: if $\Psi \approx 1$, state "B" is populated even before binding and the substrate can select it over state "A." (3) Induced fit: if $\Psi >> 1$, state "B" has to be induced by the substrate to lower its free energy upon substrate binding. Since Ψ can vary continuously from negative to positive values, scenarios between the three classic "either-or" scenarios are possible. It should also be clear that Ψ cannot always be determined from a bulk thermodynamic experiment. The free energy difference $\Delta G_B$(unbound) cannot be measured in equilibrium if $G_B$(unbound) >> $G_A$(unbound). Instead, a kinetic experiment is required that can populate state "B" under unbound conditions.

Our data provides some evidence that the Lys22Gln mutation shifts Ψ from induced fit towards conformational selection (Figure 3.12a). In the pseudo wild type Phe56Trp, fast kinetic equilibration is observed between two RNA-bound states, but tryptophan fluorescence detects no conformational relaxation of the protein by itself. The Lys22Gln mutant shows a similar equilibration between RNA-bound states, but even without RNA the protein has a phase with a similar relaxation time, but opposite amplitude. This is consistent with an equilibrium between two protein conformations in the absence of RNA binding, where both conformations have similar free energy ($\Psi \approx 1$). In that case, the tight binding site is 'selected' in the Lys22Gln mutant, but 'induced' in the pseudo wild type, as illustrated in Figure 3.12b. For Lys22Gln, a

sign switch of the amplitude indicates that the tryptophan lifetime changes in opposite directions between the two conformations.

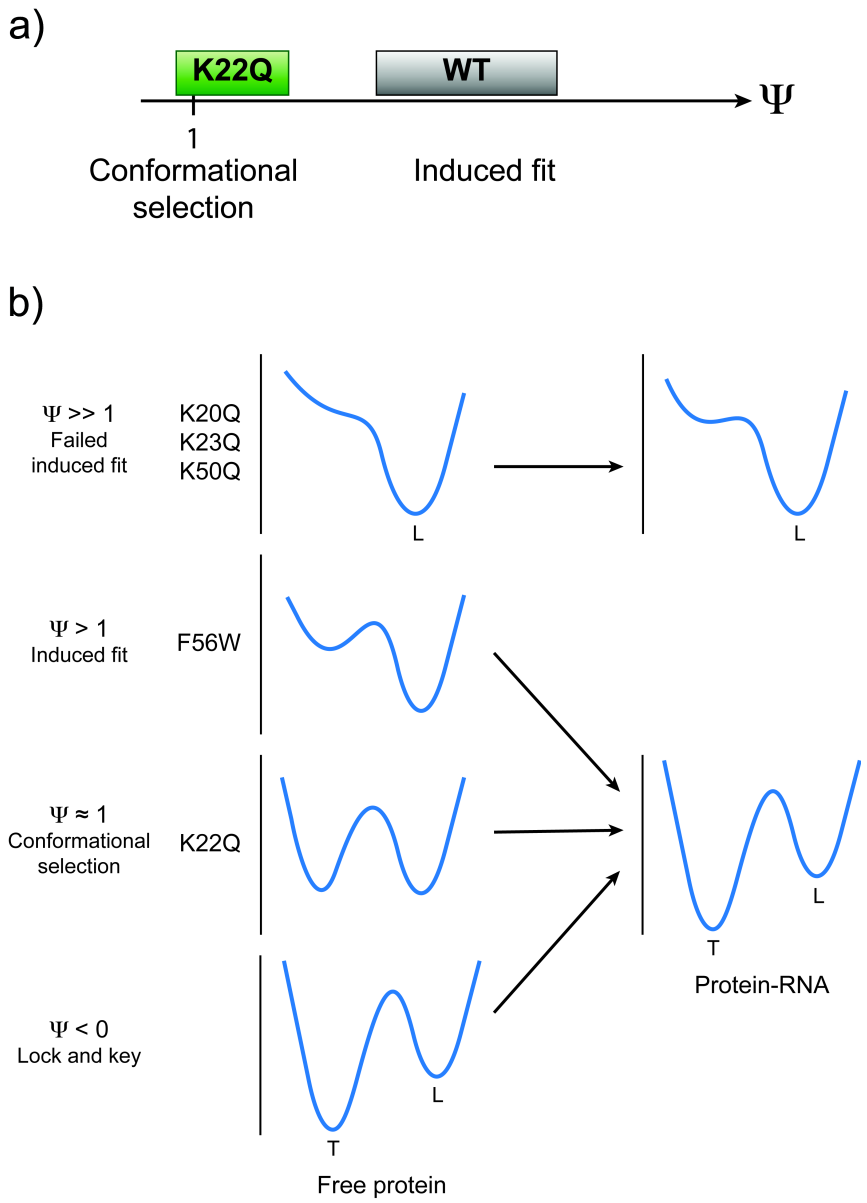



**Figure 3.12** Conformational selection and induced fit. a) Suggested position of WT and K22Q mutant in the continuum of induced fit and conformational selection. b) Energy diagram showing the continuum of induced fit and conformational selection.

Previous MD simulations have shown two substates for free U1A protein, suggesting a form of conformational selection of the U1A protein upon RNA binding, although the conversion between the two substates was not observed in the simulation. The time scale of 5 ns may not have been sufficient to reveal the conversion between the two substates. The phase observed with Lys22Gln mutant is the first experimental data that suggests some extent of conformational selection may occur in the binding process of U1A and SL2 RNA.

A recent NMR study of U2AF, a splicing factor that contains two RRMs, has shown that U1A has two conformations with and without the target polypyrimidine tract RNA. The two RRMs are oriented in two different conformations: open and closed. RNA binding shifts the equilibrium to the open form revealing a conformational selection mode of binding to RNA.[29] Together with our data, conformational selection may be rather a common theme in RNA binding proteins upon RNA binding. Therefore a combination of conformational selection and induced fit may be a more accurate description of RNA binding protein binding to its target RNA rather than pure induced fit. Future studies with other RNA binding proteins will provide a more generalized picture of the binding mechanism.

### *3.4 Conclusion*

Mutational studies of positively charged residues that were known to be important for association and complex stability of U1A and SL2 RNA support our two-step dissociation pathway and give new insights into the role of these residues in the complex. We characterized the role of each positively charged residue by combining temperature jump and stopped-flow experiments using tryptophan as a fluorescent probe. We found that depending on the position and distance between the positively charged residue and the RNA, the contribution to the

complex stability varied at each step. Residues that interact with the loop region of the RNA are crucial for the stability of the tight complex whereas a residue that contacts the stem region of the RNA is important for the stability of the loose complex. These results, with the results of previous investigations of positively charged residues, demonstrate the importance of electrostatic interactions between RBPs and RNA, and can be used to engineer proteins with tighter or looser complex populations, as well as for the design of small molecules that target one or both complexes.

### *3.5 Materials and Methods*

**Proteins expression and purification**

All mutants were generated by site-directed mutagenesis and confirmed by sequencing. $His_6$-containing mutant plasmids were transformed into *Escherichia coli* strain BL21DE3 (pLysS) competent cells. The cells were grown in LB medium, and protein expression was induced with 1 mM IPTG at $OD_{600}$=0.60. The cultures were grown for 5–6 h after induction. The cells were pelleted, resuspended in 10 mL of lysis buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 20 mM imidazole, pH 7.6) and lysed by ultrasonication. The lysate was centrifuged at 10,000 rpm and the supernatant was loaded on a 1 mL Ni-NTA column, and the protein was eluted with elution buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 250 mM imidazole, pH 7.6). Eluted protein was dialyzed against storage buffer (10 mM potassium phosphate, 50 mM KCl) and concentrated by Amicon filter MWCO 3,000. The concentration of each protein was determined using BCA assay (Pierce). Expressed proteins were confirmed by low-resolution electrospray ionization mass spectrometry and the purities of the proteins were assessed by SDS-PAGE.

**Electrophoretic mobility shift assay (EMSA)**

Varying amounts of U1A protein were incubated with 50 pM [$\gamma$-$^{32}$P] ATP-labeled RNA for 30 minutes at room temperature in a buffer containing 10 mM potassium phosphate (pH 7.4) and 1M KCl. Electrophoreses were carried out at 25 °C for 40 minutes at 350 V with 0.5X TBE as running buffer. The native gels, 8% acrylamide (42:1 = acrylamide : biacrylamide, 15 cm x 40 cm x 1.5 mm), were pre-run at 350 V for 30 minutes before loading 10 μl of reaction mixture per well. The temperature of the gel was maintained at 25 °C by a circulating water bath. Gels were visualized on a Molecular Dynamics Storm phosphorimager. Fraction RNA bound versus protein concentration was plotted and curves were fitted to the equation: Fraction bound = $1/(1+K_d/[P]_T)$ using KaleidaGraph (Synergy Software, PA).

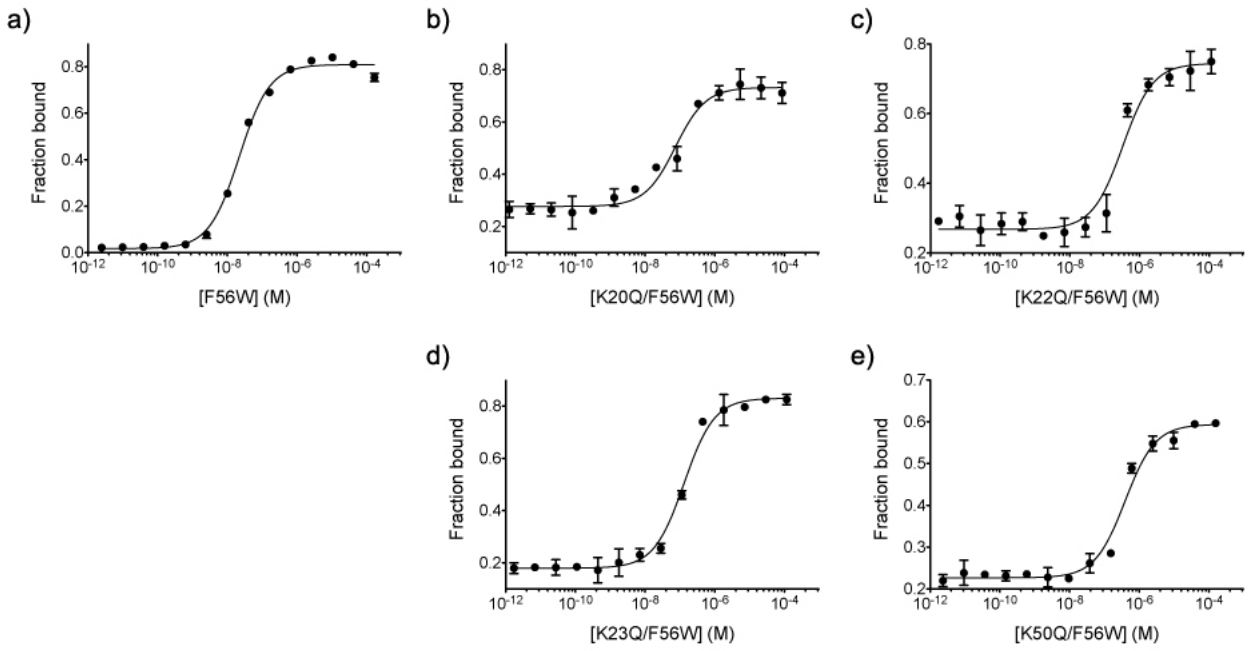



**Figure 3.13** Binding curves of U1A mutants at 200 mM KCl and 25°C. a) F56W b) K20Q/F56W c) K22Q/F56W d) K23Q/F56W e) K50Q/F56W.

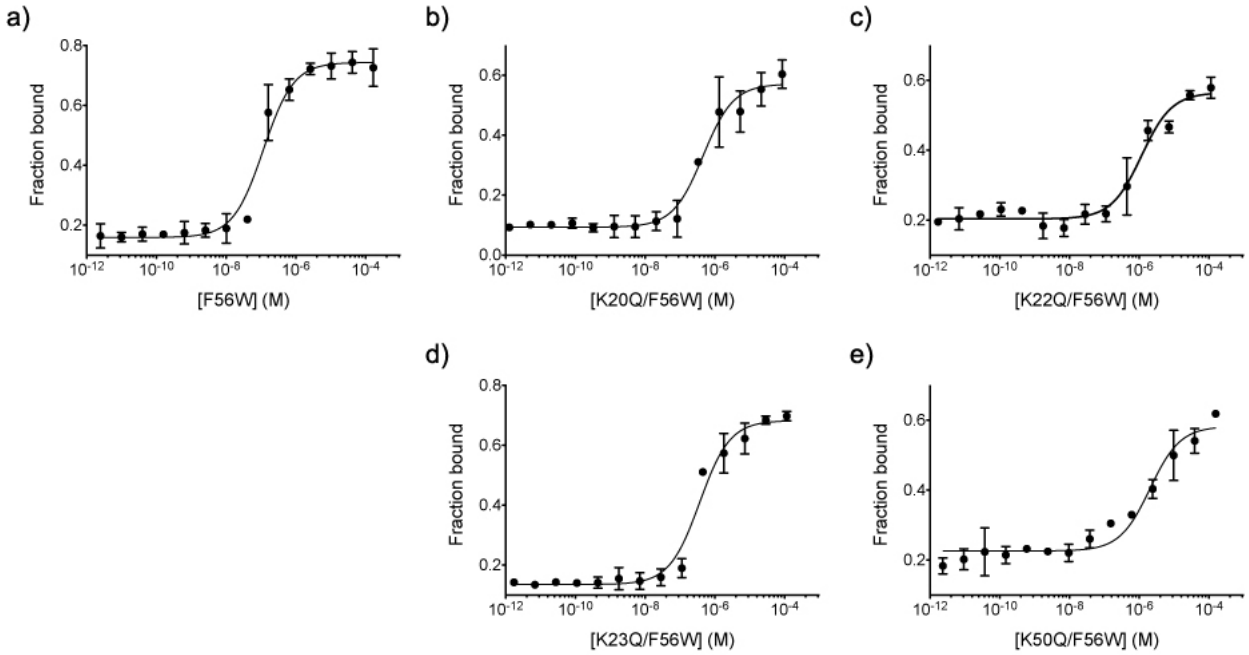



**Figure 3.14** Binding curves of U1A mutants at 200 mM KCl and 35°C. a) F56W b) K20Q/F56W c) K22Q/F56W d) K23Q/F56W e) K50Q/F56W.

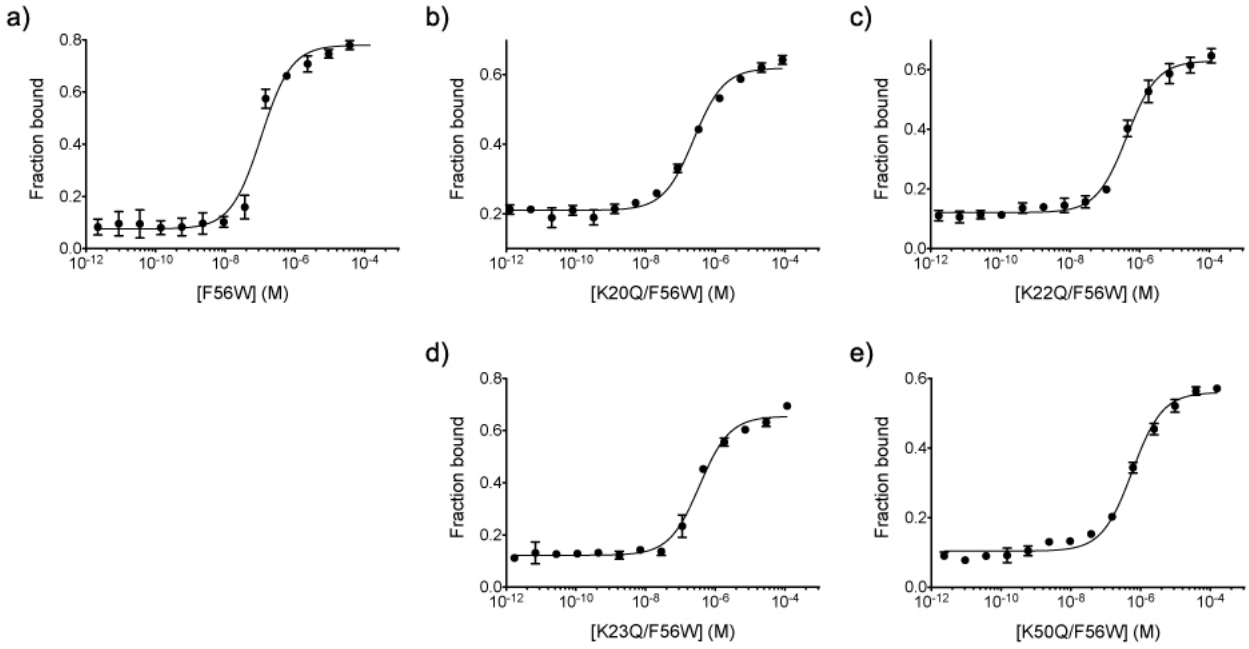



**Figure 3.15** Binding curves of U1A mutants at 1 M KCl and 25°C. a) F56W b) K20Q/F56W c) K22Q/F56W d) K23Q/F56W e) K50Q/F56W.

**Circular Dichroism (CD)**

Thermal stability of U1A mutants were measured in a J-715 spectrpolarimeter equipped with a Peltier temperature control (Jasco Inc.) at 25, 35, and 45 °C. CD spectra of 10 μM samples were prepared in sodium cacodylate buffer (10 mM cacodylic acid, 50 mM NaCl, pH 7.4) in the 200−250 nm region. Each CD spectrum in figures is an average of 50 spectra at 200 nm/min scan speed. A quartz cuvette (Starna Cells Inc.) with 200 mm path length was used.

**Laser-induced temperature jump**

Relaxation kinetics of the U1A mutants and the equilibrated U1A–SL2 RNA in binding buffer (10 mM potassium phosphate, 200 mM KCl, pH 7.4) was measured with a home-built temperature jump apparatus described elsewhere.[16] The SL2 RNA was heat-shocked before complexation to ensure correct secondary structure conformation and decrease significantly the dissociation constant ($K_d$).

Laser temperature jumps of 8-10 °C were achieved using a Surelite III Q-switched Nd:YAG laser (Continuum Inc.) Raman-shifted to 1.9 μm by passing the beam through a 1 m long tube with hydrogen gas pressurized to 300 psi. The beam was then passed through a 50 % beam splitter to allow the sample to be excited from two sides providing more uniform heating. The pre-jump equilibrium temperature was set using an automated temperature controller, model Lake Shore 330 (Lake Shore Cryotronics Inc.). The sample cell was made of fused silica tubing 3530S-100 (VitroCom) fused shut on one side. The fluorescence excitation path length was 0.3 mm. The sample was excited with a 80 MHz pulsed Ti:sapphire laser (KMLabs Inc.). The Ti:sapphire laser wavelength was 860 nm, which was frequency tripled with a third harmonic generator (CSK Optronics Inc.) to 287 nm. Tryptophan fluorescence was then guided from the

sample by an optical light-guide (Oriel Corp.), passed through a B370 band-pass filter (Hoya Corp.) and collected by a photomultiplier R7400U-03 (Hamamatsu Corp.). The signal was then recorded and digitized every 100 ps by an oscilloscope DPO7254 (Tektronix Inc.) with 2.5 GHz bandwidth. The length of the time traces was 500 μs and each trace contained many tryptophan fluorescence decays every 12.5 ns (80 MHz). The temperature jump was set to occur 153.75 μs after the oscilloscope was triggered to start data collection to provide a pre-jump baseline. The fluorescence decay peak signal was usually 10-40 mV.

**T-jump Fluorescence lifetime decays analysis**

Relaxation kinetics data were analyzed using MATLAB (Mathworks Inc.). Time traces were first binned into the intervals of 100 decays, which corresponds to 1.25 μs. We analyzed the data by fitting each averaged fluorescence decay $f$ as a linear combination of two fluorescence decay profiles, $f = a_1(t) f_1 + a_2(t) f_2$, where $f_1$ is the fluorescence profile before the temperature jump, and $f_2$ represents the equilibrated fluorescence profile after the temperature jump. The parameter $\chi$ is defined by $c(t) = a_1 / (a_1 + a_2)$ to track the progression of the fluorescence decay profile (i.e. change in tryptophan lifetime) from before the temperature jump ($\chi_1$=1), to when equilibrium has been reached again ($\chi_1$=0). The time traces were then fitted to a single exponential function starting with point zero where the jump occurred.

**Stopped-flow**

U1A-SL2 dissociation kinetics was monitored using fast mixing measurements performed using a SX-20 stopped-flow spectrometer (Applied Photophysics). The tryptophan was excited at 280 nm and fluorescence emission was monitored through a 350 nm interference

filter (Applied Photophysics) with 2 mm entrance and exit slit width. The fast mixing time is approximately 3 ms giving a dead time of ~1 ms. Fluorescence scans were collected in 10 s data files with 1000 data points. During dissociation studies a syringe with 0.5 μM U1A–RNA complex solution in binding buffer (10 mM potassium phosphate, 200 mM KCl, pH 7.4) was fast mixed with phosphate buffer (10 mM potassium phosphate, 1.8 M KCl, pH 7.4) from another syringe. A final concentration of 1 M KCl was required to obtain an optimal dissociation fluorescence signal. For each sample, at least seven individual scans were average to give one data set. Three independent averages were use for data analysis.

### *3.6 References*


1 Blackshear, P. J.; Lai, W. S.; Kennington, E. A.; Brewer, G.; Wilson, G. M.; Guan, X.; Zhou, P. Characteristics of the interaction of a synthetic human tristetraprolin tandem zinc finger peptide with AU-rich element-containing RNA substrates, *J Biol Chem* **2003**, *278*, 19947.

2 Brewer, B. Y.; Ballin, J. D.; Fialcowitz-White, E. J.; Blackshear, P. J.; Wilson, G. M. Substrate dependence of conformational changes in the RNA-binding domain of tristetraprolin assessed by fluorescence spectroscopy of tryptophan mutants, *Biochemistry* **2006**, *45*, 13807.

3 Perona, J. J.; Hou, Y. M. Indirect readout of tRNA for aminoacylation, *Biochemistry* **2007**, *46*, 10419.

4 Mittermaier, A.; Varani, L.; Muhandiram, D. R.; Kay, L. E.; Varani, G. Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein, *J Mol Biol* **1999**, *294*, 967.

5 Jean, J. M.; Clerte, C.; Hall, K. B. Global and local dynamics of the human U1A protein determined by tryptophan fluorescence, *Protein Sci* **1999**, *8*, 2110.

6 Avis, J. M.; Allain, F. H.; Howe, P. W.; Varani, G.; Nagai, K.; Neuhaus, D. Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding, *J Mol Biol* **1996**, *257*, 398.

7 Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C. H.; Nagai, K. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature* **1994**, *372*, 432.

8 Showalter, S. A.; Hall, K. B. Correlated motions in the U1 snRNA stem/loop 2: U1A RBD1 complex, *Biophysical Journal* **2005**, *89*, 2046.

9 Blakaj, D. M.; McConnell, K. J.; Beveridge, D. L.; Baranger, A. M. Molecular dynamics and thermodynamics of protein-RNA interactions: Mutation of a conserved aromatic residue modifies stacking interactions and structural adaptation in the U1A-stem loop 2 RNA complex, *J Am Chem Soc* **2001**, *123*, 2548.

10 Reyes, C. M.; Kollman, P. A. Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change, *J Mol Biol* **2000**, *297*, 1145.

11 Hall, K. B. Interaction of Rna Hairpins with the Human U1a N-Terminal RNA-Binding Domain, *Biochemistry* **1994**, *33*, 10076.

12 Katsamba, P. S.; Myszka, D. G.; Laird-Offringa, I. A. Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA, *J Biol Chem* **2001**, *276*, 21476.

13 Katsamba, P. S.; Bayramyan, M.; Haworth, I. S.; Myszka, D. G.; Laird-Offringa, I. A. Complex role of the beta(2)-beta(3) loop in the interaction of U1A with U1 hairpin II RNA, *J Biol Chem* **2002**, *277*, 33267.

14 Law, M. J.; Chambers, E. J.; Katsamba, P. S.; Haworth, I. S.; Laird-Offringa, I. A. Kinetic analysis of the role of the tyrosine 13, phenylalanine 56 and glutamine 54 network in the U1A/U1 hairpin II interaction, *Nucleic Acids Res* **2005**, *33*, 2917.

15 Law, M. J.; Linde, M. E.; Chambers, E. J.; Oubridge, C.; Katsamba, P. S.; Nilsson, L.; Haworth, I. S.; Laird-Offringa, I. A. The role of positively charged amino acids and electrostatic interactions in the complex of U1A protein and U1 hairpin II RNA, *Nucleic Acids Res* **2006**, *34*, 275.

16 Anunciado, D.; Dhar, A.; Gruebele, M.; Baranger, A. M. Multistep kinetics of the U1A-SL2 RNA complex dissociation, *J Mol Biol* **2011**, *408*, 896.

17 Garcia-Garcia, C.; Draper, D. E. Electrostatic interactions in a peptide--RNA complex, *J Mol Biol* **2003**, *331*, 75.

18 Shazman, S.; Mandel-Gutfreund, Y. Classifying RNA-binding proteins based on electrostatic properties, *PLoS Comput Biol* **2008**, *4*, e1000146.

19 Tworowski, D.; Safro, M. The long-range electrostatic interactions control tRNA-aminoacyl-tRNA synthetase complex formation, *Protein Sci* **2003**, *12*, 1247.

20 Tworowski, D.; Feldman, A. V.; Safro, M. G. Electrostatic potential of aminoacyl-tRNA synthetase navigates tRNA on its pathway to the binding site, *J Mol Biol* **2005**, *350*, 866.

21 Lu, J. R.; Hall, K. B. An Rbd That Does Not Bind Rna - Nmr Secondary Structure Determination and Biochemical-Properties of the C-Terminal Rna-Binding Domain from the Human U1a Protein, *J Mol Biol* **1995**, *247*, 739.

22 Kurihara, Y.; Nagata, T.; Imai, T.; Hiwatashi, A.; Horiuchi, M.; Sakakibara, S.; Katahira, M.; Okano, H.; Uesugi, S. Structural properties and RNA-binding activities of two RNA recognition motifs of a mouse neural RNA-binding protein, mouse-Musashi-1, *Gene* **1997**, *186*, 21.

23 Miyanoiri, Y.; Kobayashi, H.; Imai, T.; Watanabe, M.; Nagata, T.; Uesugi, S.; Okano, H.; Katahira, M. Origin of higher affinity to RNA of the N-terminal RNA-binding domain than that of the C-terminal one of a mouse neural protein, musashi1, as revealed by comparison of their structures, modes of interaction, surface electrostatic potentials, and backbone dynamics, *J Biol Chem* **2003**, *278*, 41309.

24 Shiels, J. C.; Tuite, J. B.; Nolan, S. J.; Baranger, A. M. Investigation of a conserved stacking interaction in target site recognition by the U1A protein, *Nucleic Acids Res* **2002**, *30*, 550.

25 Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem* **2004**, *25*, 1605.

26 Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions, *Physica* **1940**, *7*, 284.

27 Changeux, J. P.; Edelstein, S. Conformational selection or induced fit? 50 years of debate resolved, *F1000 Biol Rep* **2011**, *3*, 19.

28 Csermely, P.; Palotai, R.; Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events, *Trends biochem sci* **2010**, *35*, 539.

29 Mackereth, C. D.; Madl, T.; Bonnal, S.; Simon, B.; Zanier, K.; Gasch, A.; Rybin, V.; Valcarcel, J.; Sattler, M. Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF, *Nature* **2011**, *475*, 408.

# CHAPTER 4

# Bipartite Cysteine Display Probes for the Detection of RNA Splice Variants

## *4.1 Introduction*

RNA splicing is an essential process for cellular functions, and disruption of precise RNA splicing is often involved in diseases. Therefore, there is a critical need for modulators of RNA splicing that can target specific splicing factors. This chapter will describe a novel method to detect different splice products, which could lead to an assay to screen for small molecule activators or repressors of RNA splicing.

### *4.1.1 RNA Splicing and diseases*

Human genes express complex pre-mRNAs that contain exons that will make up the mRNA product and non-coding introns that interrupt the exons.[1] Due to this nature of pre-mRNA, RNA splicing is necessary. The introns of the RNAs are spliced out and the exons are joined together to create the reading frame for translation. In addition to constitutive splicing, approximately 95% of human genes are alternatively spliced, which allows a single gene to express multiple proteins.[2] Most of alternative splicing occurs within open reading frames (ORFs), expanding the human proteome. In other cases, alternative splicing occurs within untranslated regions and affect *cis*-acting elements that control mRNA stability, translation efficiency, and mRNA localization. Alternative splicing may also introduce premature termination codons, which in turn undergo nonsense-mediated decay. RNA splicing is accomplished by the spliceosome, a macromolecular ribonucleoprotein complex that assembles

on the pre-mRNA. The spliceosome is composed of five small nuclear ribonucleoprotein (snRNP) particles (U1, U2, U4, U5 and U6) and a large number of additional proteins.

*Cis*-acting elements direct the spliceosome to recognize the *bona fide* exons, remove introns, and serve as binding sites for splicing factors that regulate alternative splicing. These elements are also called 'splicing code', which include the consensus splice site sequences and auxiliary elements known as exon and intron splicing enhancers (ESEs and ISEs respectively) and silencers (ESSs and ISSs respectively).[1] The splicing code regulates both constitutive and alternative splicing. The secondary structure of the RNA also plays a role in splicing regulation by sequestering sequence elements or bringing two elements into close proximity.[3] Furthermore, the rate of transcription elongation has been shown to have an effect on alternative splicing by accelerating or delaying the synthesis of competing splice sites or regulatory elements.[4]

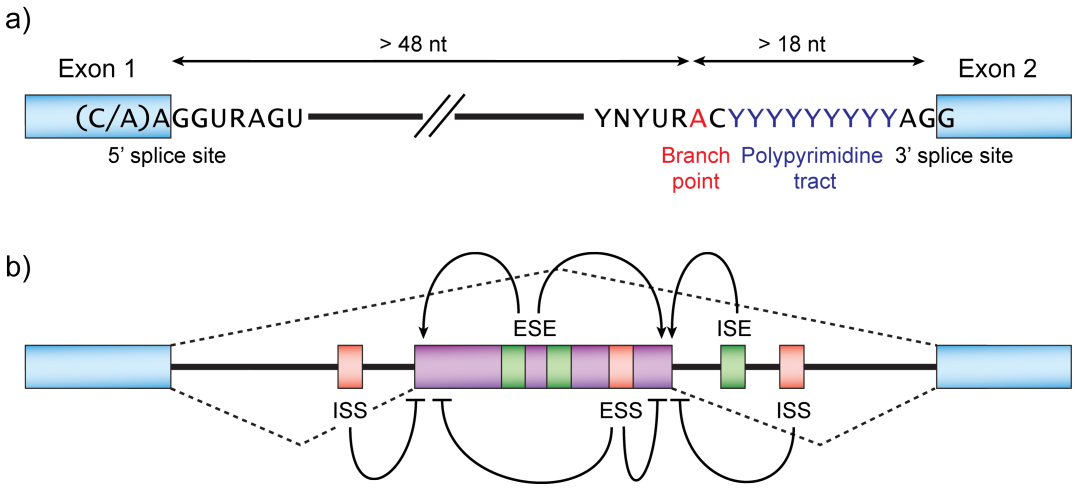



**Figure 4.1** Splicing codes.[1] a) Consensus splice site sequences. b) Exon and intron splicing enhancers and silencers.

*Trans*-acting elements include SR proteins and hnRNP proteins that bind to the enhancer and silencer sites, respectively. These proteins communicate with other proteins to either recruit

or repress the activity of the spliceosome. In addition, RNA polymerase II and transcription factors can directly or indirectly interact with splicing factors and either activate or repress alternative splicing. Together, RNA splicing is controlled by *cis*-acting elements, many *trans*-acting factors and other factors such as RNA structure, rate of transcription, RNA polymerase II, and transcription factors, which make this process highly versatile, complex, and dynamic.

Mutations that disrupt the splicing code or *trans*-acting elements can directly cause disease, modify the severity of the disease phenotype, or be linked with disease susceptibility.[5] Mutations in the introns as well as the exons, including those that are silent at the translational level, may induce exon skipping, create new exon boundaries, activate cryptic splice sites, or create pseudoexons. The effects of these mutations are more direct and cause aberrant splicing of one gene. However mutations in the *trans*-acting elements or proteins that are involved in the spliceosome may cause mis-splicing of mutiple genes.[5] Diseases that are caused by aberrant RNA splicing include Duchene muscular dystrophy, cystic fibrosis, spinal muscular atrophy, retinitis pigmentosa, β-thalassaemia, and several types of cancer, to name just a few.[5,6]

Therefore, small molecule modulators of RNA splicing can be used as therapeutic agents and as tools to understand the complicated splicing regulatory mechanism. Efficient detection of RNA splice variants may aid the identification of small molecules that modulate RNA splicing and accelerate biochemical studies of RNA splicing.

*4.1.2 In vitro and in-cell RNA splicing assays.*

*In vitro* splicing assays have led to the discovery of many fundamental features and components of RNA splicing.[7] Although the standard *in vitro* splicing system is limited by the fact that RNA splicing events are separated from transcription and translation, the versatility of

the assay has made it a powerful tool to dissect the mechanism of RNA splicing. The major advantage of the *in vitro* splicing assay is its tremendous flexibility. Biochemical approaches that modify the reaction conditions may lead to isolation, accumulation, and characterization of intermediates, which give valuable insight into how the splicing is accomplished as well as what regulatory factors are involved. A typical *in vitro* splicing assay is carried out in nuclear extract, which contains the spliceosome and basic *trans*-acting factors for splicing. The pre-mRNA of interest is prepared by *in vitro* transcription using radioactive nucleoside triphosphate–usually uridine triphosphate (UTP). The spliced products are separated and visualized by polyacrylamide gels and phosphorimagers analysis.[7]

In-cell splicing assays are crucial to understand the splicing mechansim in the natural context. It has been found that the rate of intron removal is significantly slower in *in vitro* assays. And many RNA binding proteins that are involved in splicing, shuttle between the nucleus and cytoplasm. Therefore in cell studies, which include the cytoplasm and other gene expression steps into account, provide a more accurate and complete picture of the splicing process. Assays involve the isolation of RNA from the whole cell extract and measurement of mRNA by methods such as RT-PCR.

*4.1.3 High throughput screening methods*

Small molecules provide a valuable tool to understand RNA splicing as well as other cellular processes. For example, Spliceostatin A, which inhibits splicing and nuclear retention of pre-mRNA, has been used to reveal the functions of U1 snRNP's that are independent of splicing.[8,9] Other small molecules that modify splicing patterns have been identified using several different strategies, including RT-PCR, reporters producing GFP, and luciferase.[10-13] The

reporter constructs are designed by introducing an intron in the open reading frame so that only the spliced RNA will produce the reporter protein. Each of these assays has limitations in a high-throughput setting. RT-PCR cannot be easily scaled up and is costly. The reporter systems assume the mRNA level is proportional to the expressed protein level, which may not be true in many cases.[14] Another disadvantage of these methods is that it is hard to distinguish between compounds that modulate RNA splicing and those that affect transcription, or translation.[14,15] More sophisticated reporter constructs that overcome some of the limitations have been reported. However these reports recommend using RT-PCR or western blots to confirm that the obtained signal is proportional to the actual RNA splicing level.[15]

*In vitro* splicing assays may have some advantages over cell-based high throughput screening assays. They do not involve the expertise in gene manipulation that is required to make the reporter construct and the reaction conditions can be easily modified. However the detection step, which involves the use of radiolabeled RNAs and polyacrylamide gels, is the limiting factor for high-throughput *in vitro* splicing assays.

### *4.1.4 Current methods to detect DNA and RNA*

Although there are a limited number of reports that directly aim for the detection of RNA splice variants, there are several nucleic acid detection methods that can be adapted for RNA splicing assays. Hybridization of nucleic acids has been utilized to develop such methods. Among those methods the most widely recognized is molecular beacons. A molecular beacon is an antisense oligonucleotide probe that contains a fluorophore at one end and a quencher at the other end (Figure 4.2). The probe forms a stem loop in the absence of the target nucleic acid, which brings the fluorophore and the quencher in close proximity.[16] Encounter with the target nucleic acid separates the fluorophore from the quencher, which leads to an increase of

fluorescence. Since the introduction of molecular beacons in 1996, many efforts have been made to improve their design and extend their applications.[16] Currently, a large family of molecular beacons are used in many biological studies, including genetic screening, biosensor development, biochip construction, the detection of singlenucleotide polymorphisms (SNPs), and messenger-RNA (mRNA) monitoring in living cells.[16]

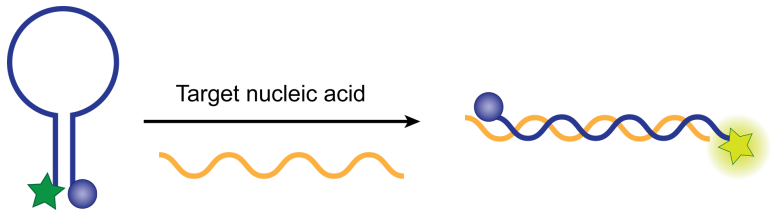



**Figure 4.2** Molecular beacon. Blue circle is the quencher and green star is the fluorophore.

Other antisense oligonucleotide based probes include quenched autoligation (QUAL) probes, quenched Staudinger-triggered α-azido ether release (Q-STAR) probes, and junction probes (see Figure 4.3 and 4.4 for the design of each probes).[17-20] QUAL probes and Q-STAR probes rely on chemical reactions ($S_N2$ reaction and Staudinger reaction, respectively) that are templated by the target nucleic acid. QUAL probes are limited by slow reaction rate and undesired reactions with endogenous nucleophiles.[17,19] Thus, Q-STAR probes, which utilize an orthogonal Staudinger reaction, were developed.[19] Junction probe methods utilize two probes that do not hybridize, but can form a triplex with the target nucleic. When the trimolecular nucleic acid forms, it presents a cognate restriction endonuclease site that is cleaved. This cleavage reaction releases the quencher, which leads to fluorescence.[20,21] The dissociation of the shortened probes and association of new probes amplifies the signal. Although less simple compared to molecular beacons, the amplified signal is a significant advantage of QUAL, Q-

STAR and the junction probes. Further optimization may lead to promising probes that can be widely used.

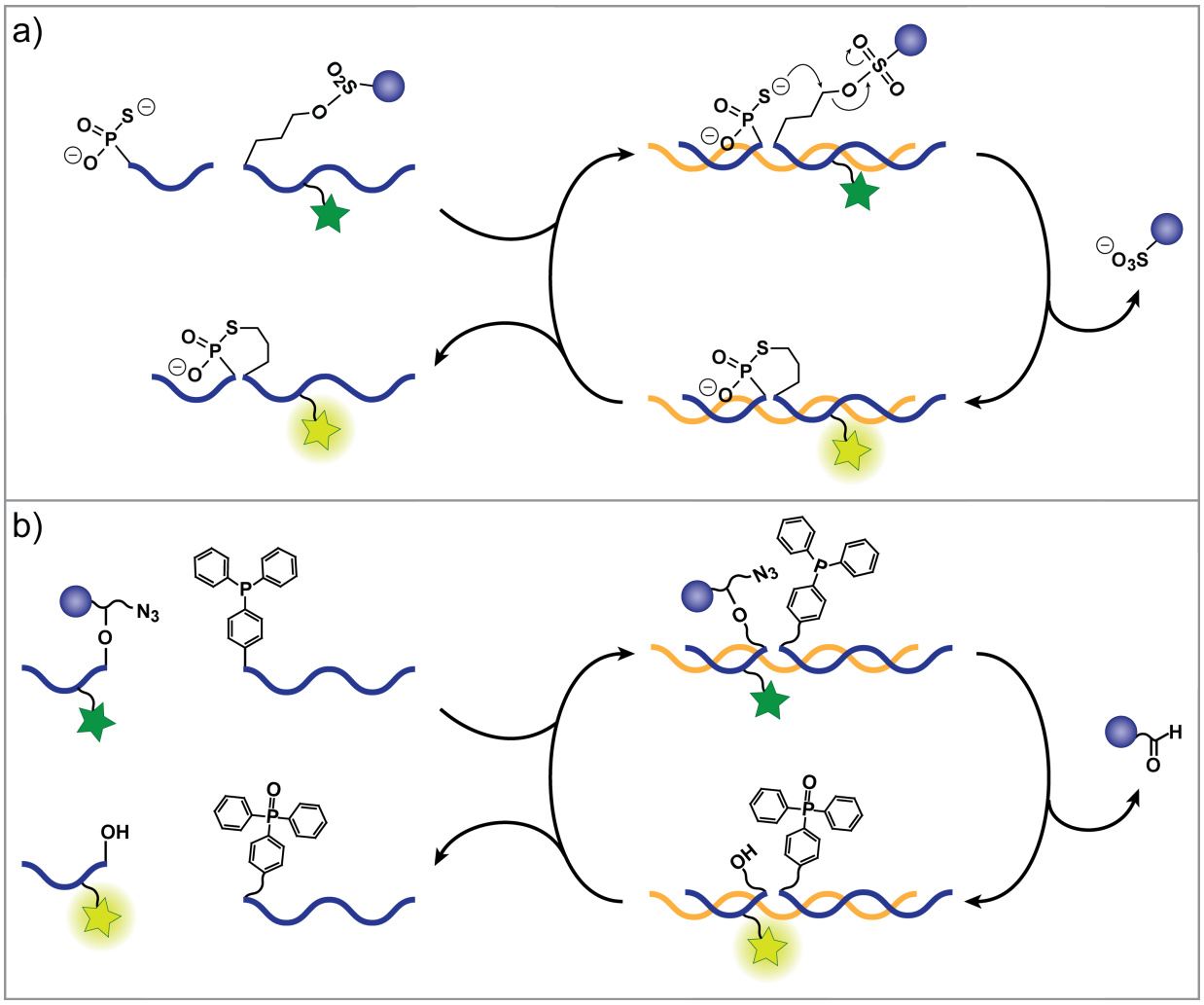



**Figure 4.3** Nucleic acid detection probes developed by Kool *et al.* a) QUAL probes. b) Q-STAR probes. Blue circle is the quencher and green star is the fluorophore.

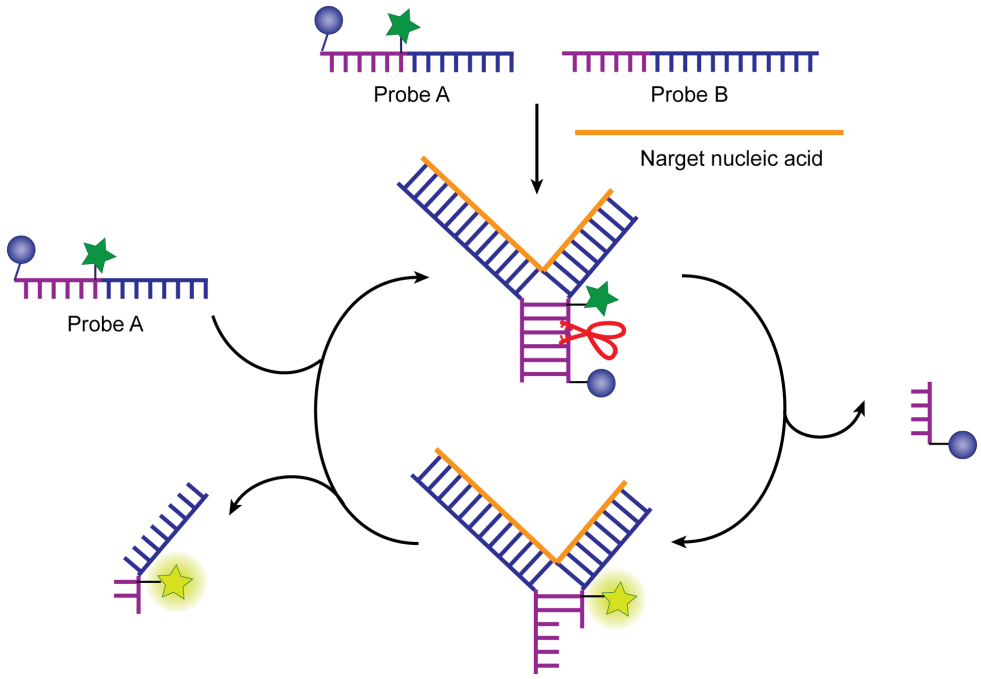



**Figure 4.4** Junction probes. Red scissors indicate a restriction enzyme that is added to the solution.

*4.1.5 Biarsenical dyes and tetracysteine complex*

In 1998, Tsien and co-workers introduced two cell permeable fluorogenic biarsenical dyes–4,5-bis(1,3,2-dithiarsolan-2-yl)fluorescein (FlAsH-$EDT_2$) and 4,5-bis(1,3,2-dithiarsolan-2-yl)resorufin (ReAsH-$EDT_2$)–that bind to tetracysteine motifs for fluorescent protein labeling (Figure 4.5a and b).[22] The biarsenical dyes are not fluorescent in the 1,2-ethanedithiol (EDT)-bound form. When the dyes are incubated with a tetracysteine peptide that is separated by any two amino acids (Cys-Cys-X-X-Cys-Cys), EDT is exchanged for the tetracysteine motif and the complex fluoresces. The biarsenical dye-tetracysteine complex is more than 50,000 times more fluorescent than FlAsH-$EDT_2$.[22] Tsien and co-workers hypothesized that the small size of EDT permits the rotation of the aryl-arsenic bond, which quenches the fluorescence by vibrational deactivation or photoinduced electron transfer. In contrast, the aryl-arsenic bond of peptide

complex is more rigid, hindering the conjugation of the arsenic lone pair electrons with the fluorescein or resorufin orbitals, which allows the complex to fluoresce.[22] The binding affinity of the FlAsH molecule and the tetracysteine peptide ranges from low pM to subnanomolar depending on the binding conditions and the identity of the amino acids in between the two dicysteine moieties.[23] The sequence of FLNCCPGCCMEP was reported to have the lowest binding affinity for the FlAsH molecule and the highest quantum yield.[23] This system was successfully used to fluorescently label proteins with a tetracysteine tag *in vitro* and in live cells.[22,24,25]

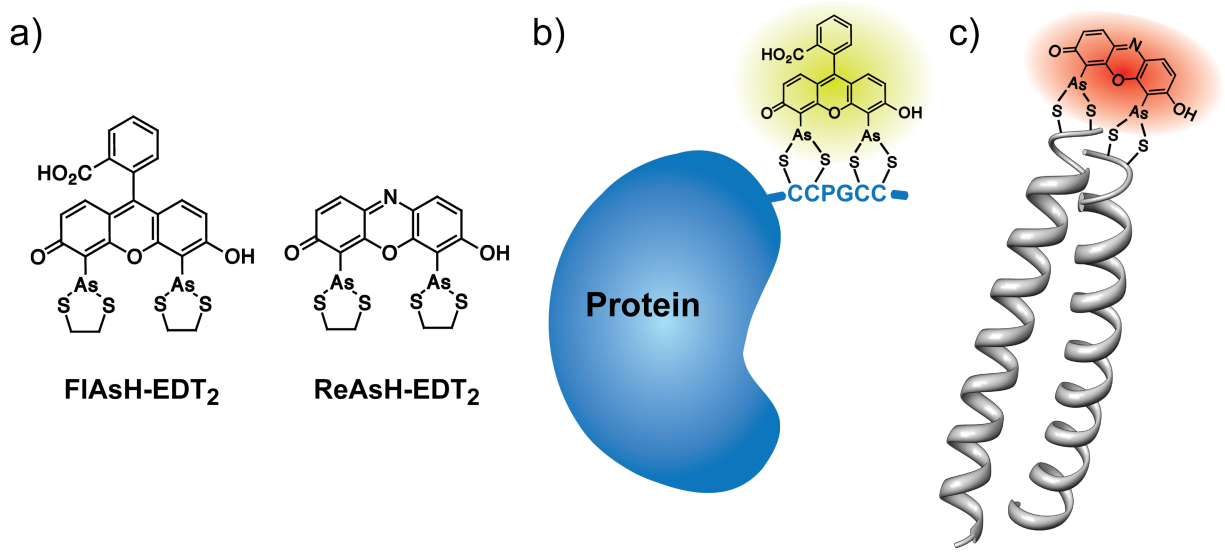



**Figure 4.5** a) Chemical structures of FlAsH-$EDT_2$ and ReAsH-$EDT_2$. b) Fluorescent labeling of protein that contains tetracysteine motif. c) Detection of peptide-peptide interaction using bipartite cysteine display.

More recently, this system was modified to detect protein-protein interactions and conformational changes of proteins. Two peptides that contain dicysteine moieties at the end of each peptide were synthesized. The peptide-peptide interaction brings the two bicysteine

moieties together forming the binding site for FlAsH molecule (Figure 4.5c).[26] With the same split tetracysteine strategy, conformational change of a protein was also detected.[27] The site to incorporate dicysteine motif was carefully selected and the fluorescence was monitored with the native protein and the denatured form of the protein. Only the native protein that was folded correctly showed fluorescence.[27]

In this study, we have designed and analyzed a novel set of probes to detect RNA splice variants based on FlAsH and split tetracysteine motif. Our results show that biarsenical dyes can bind to two dicysteine moieties that are brought together by complimentary oligonucleotides. The probes were able to detect nanomolar concentration of DNA and RNA transcripts suggesting the possibility of using these probes for a high throughput RNA splicing assay.

## *4.2 Results & Discussion*

### *4.2.1 Design of the Probes*

Biarsenical dyes 4,5-bis(1,3,2-dithiarsolan-2-yl)fluorescein (FlAsH-$EDT_2$) and 4,5-bis(1,3,2-dithiarsolan-2-yl)resorufin (ReAsH-$EDT_2$) fluoresce upon binding to tetracysteine residues. This method has been used to fluorescently label proteins and track them in living cells. Recently the Schepartz group[26] and the Gierasch group[27] reported a split tetracysteine strategy to study polypeptide and protein conformation. They added two Cys pairs at the termini of a peptide or incorporated Cys residues within the β-sheet to create FlAsH binding sites when the peptide or protein was folded.

Successful reconstitution of dye binding sites in the experiment described above suggested to us that this strategy could be used for the detection of RNA splicing following the

design shown in Figure 4.6. Thus, oligonucleotides that were complementary to two different exons were designed. Dicysteine peptides ere attached to the 3' or 5'end of the oligonucleotides such that the dicysteine groups would face each other when the oligonucleotides bound the target RNA splice variant. Addition of FlAsH-$EDT_2$ molecule should allow the detection of alternative splicing. Compared to current methods such as RT-PCR and PAGE analysis with radiolabeled RNA, this method should be a more rapid and direct way to monitor splicing. In addition, this method could be an efficient high throughput screening method for detecting small molecules that activate or repress alternative splicing.

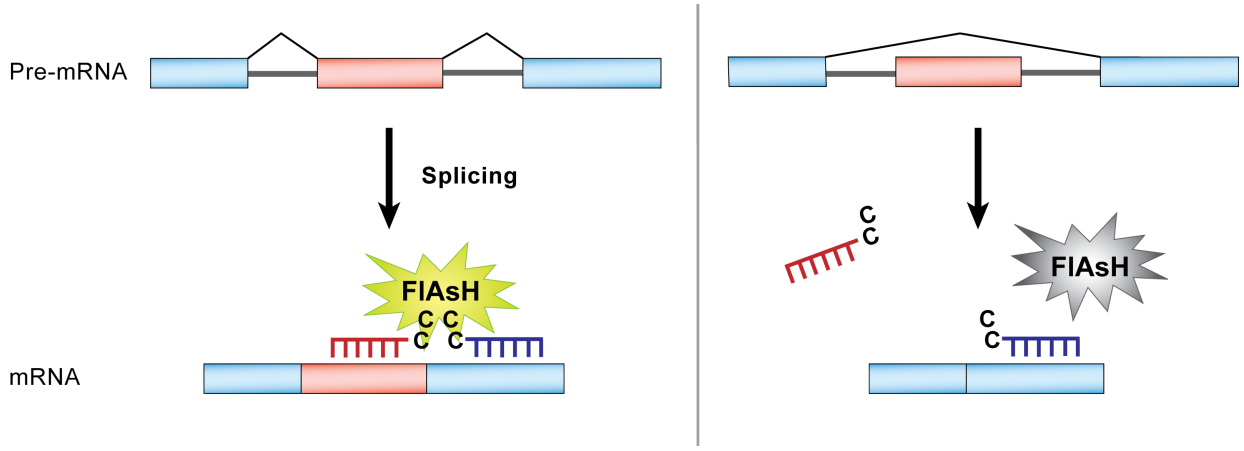



**Figure 4.6** Detection of RNA splice variants using FlAsH and dicysteine conjugated antisense oligonucleotide probes.

For the splicing system, the human survival motor neuron (*SMN*) gene was chosen. Pre-mRNA splicing of *SMN1* includes exon 7, which is translated into a functional SMN protein. An extra copy of *SMN* gene, designated *SMN2* has a critical C to T transition within the exonic splicing enhancer site. This transition abrogates the binding of *trans*-acting splicing factor SF2/ASF, which results in an unstable exon 7-skipped protein isoform (Figure 4.7).[28,29] Spinal

muscular atrophy is caused by the loss of function of both copies of *SMN1* genes.[30-32] Therefore this system has been studied intensively to restore the ability of *SMN2* gene to include exon 7 to form a functional SMN protein.[33] Quantitative detection of these two isoforms would facilitate the efficiency of these studies.

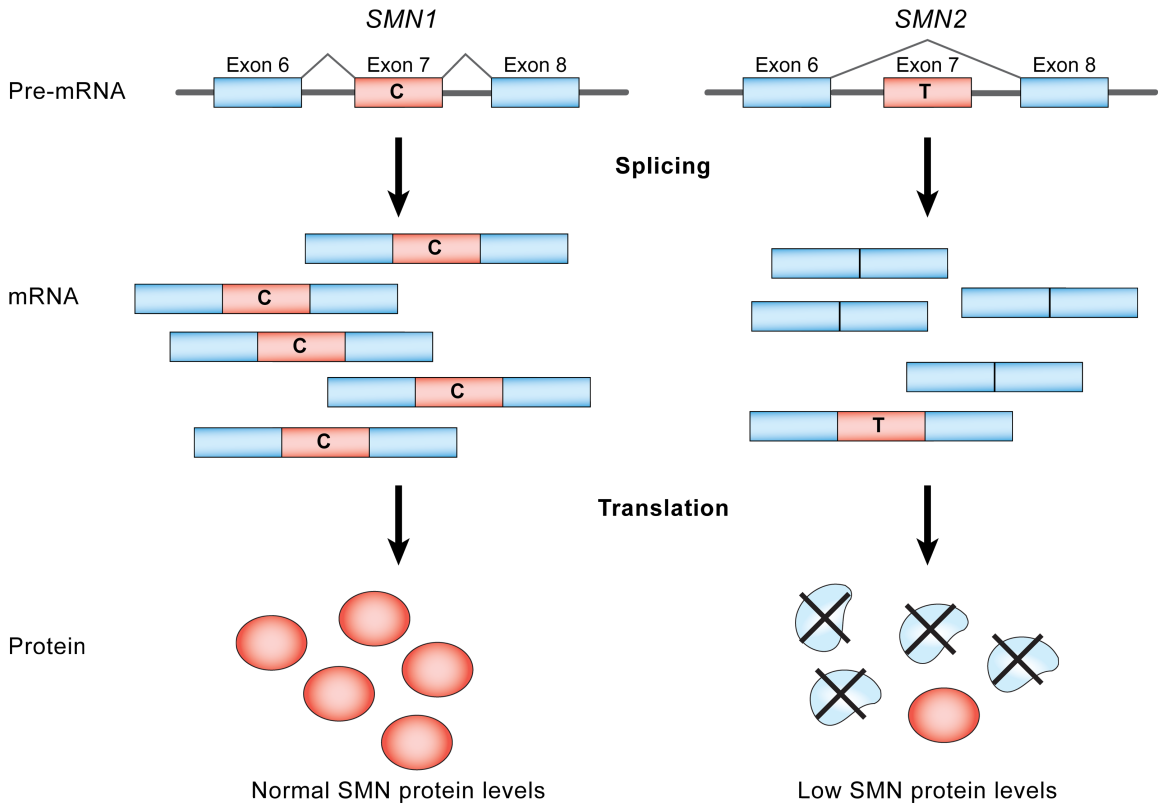



**Figure 4.7** Expression of *SMN1* and *SMN2* genes. *SMN2* gene has a T instead of C in exon 7, which results in a low level of SMN proteins.[29]

For the detection of two different splice variants of *SMN* RNA, two antisense oligonucleotide probes that are each complimentary to exon 7 and exon 8 were designed (Figure 4.8). When exon 7 is included in the splice variant, both probes may bind to the transcript positioning the two dicysteine moieties in proximity. Then FlAsH-$EDT_2$ should bind to the two dicysteine moieties, which will result in fluorescence. On the other hand, when exon 7 is

excluded, the exon 7 probe cannot be in proximity of the exon 8 probe therefore, resulting in no fluorescence. Previous studies have shown that the two dicysteine moieties need to be within 7Å for the FlAsH molecule to effectively bind the split tetracysteine.[26] Therefore, a maleimide linker was selected due to its short length and ease of synthesis.[34] Based on computational analysis with the program Molecular Operating Environment (MOE), the two probes were designed so that there would be no space between the two probes when annealed to the target RNA. This allowed the two dicysteine peptides to be positioned within 7Å of each other.

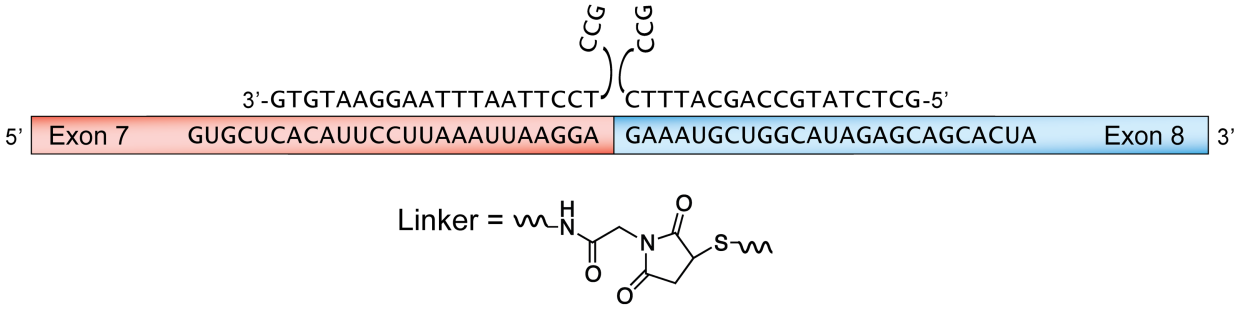
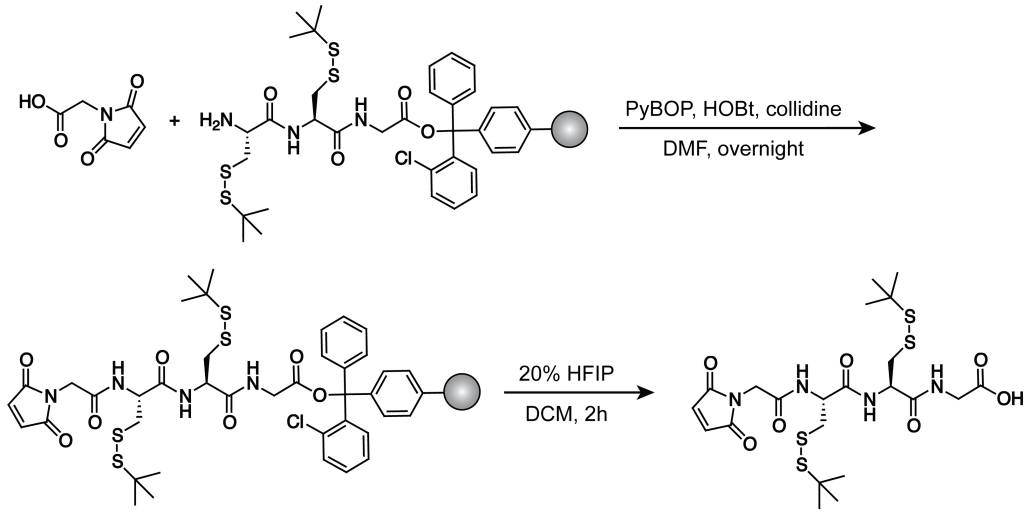



**Figure 4.8** Sequences of the two probes and the chemical structure of the linker.

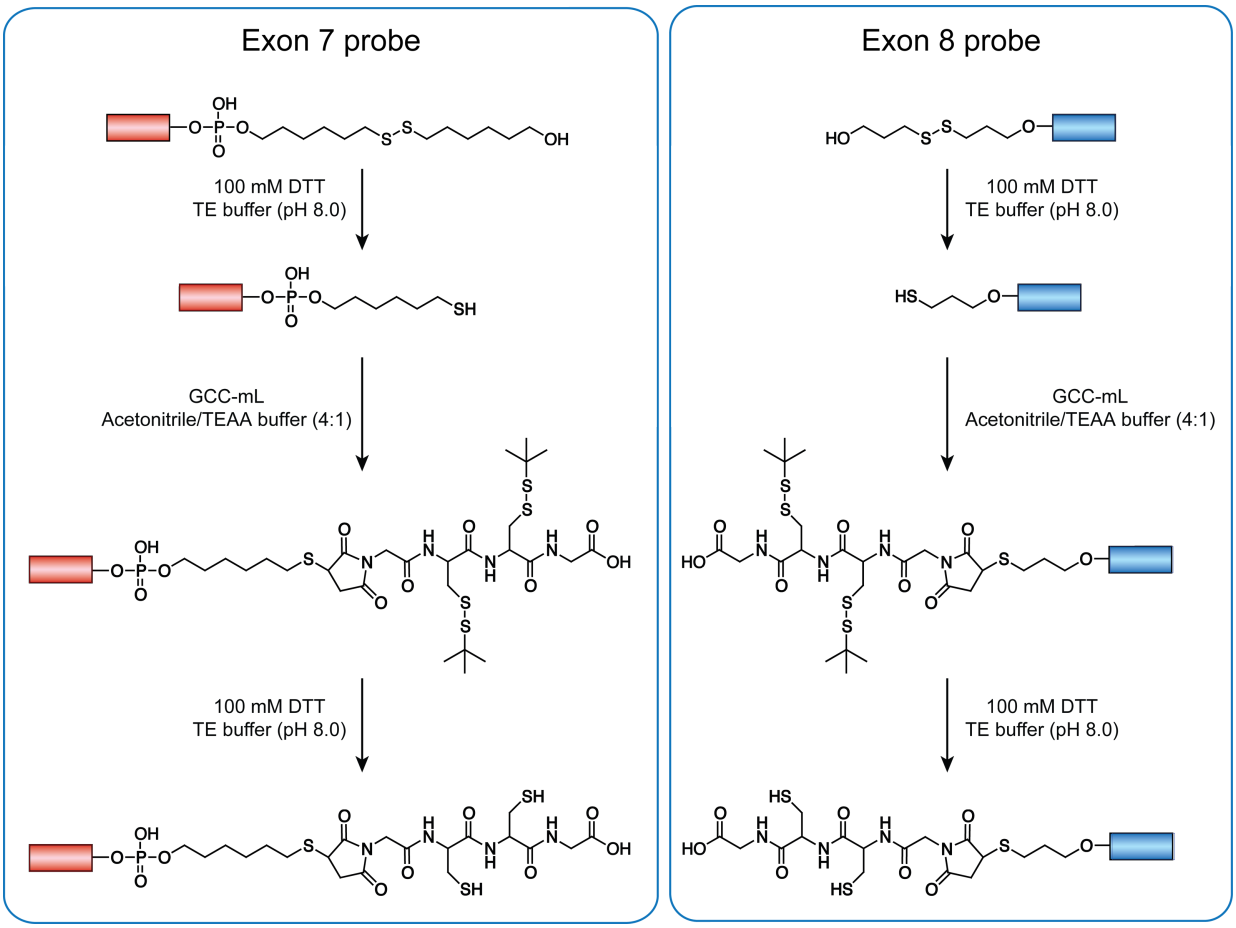
**Scheme 4.1** Synthesis of GCC-linker.



*4.2.2 Synthesis of the Probes*

The maleimide containing dicysteine peptide was prepared on solid support using standard coupling reagents (Scheme 4.1). An orthogonal protecting group of the cysteine residues, which is not cleaved during the resin cleavage is critical since the maleimide reacts with thiol groups. *Tert*-thiol butyl group suited this requirement and can be easily deprotected using dithiothreitol (DTT). The thiol groups of DNA antisense oligonucleotides were deprotected using DTT and allowed to react with the maleimide groups of the dicysteine peptides. The final product was obtained by deprotecting the cysteines (Scheme 4.2).

**Scheme 4.2** Synthesis of both probes.



*4.2.3 Detection Limit of Biarsenical Tetracysteine and FlAsH-EDT$_2$ System*

For RNA splicing assays, a high detection limit is required.[7] A typical FlAsH, tetracysteine tagged protein experiment uses low concentrations of FlAsH-EDT$_2$ and an excess of the protein. This is optimal for the detection of protein since the binding affinity of FlAsH and the peptide tag is in the low picomolar range and the low concentration of FlAsH reduces the background fluorescence of FlAsH-EDT$_2$. However the bipartite cysteine display experiments have shown that the binding affinity between the split tetracysteine motif and the FlAsH molecule is in the low micromolar range.[26,27] Therefore, in order to detect a low concentration of RNA splice variants and considering the fact that the binding affinity of the split tetracysteine motif is lower than that of intact tetracysteine motif, a higher FlAsH-EDT$_2$ concentration is required. In addition to the FlAsH-EDT$_2$ concentration, EDT concentration is an important factor to obtain good signal to noise ratio since the equilibrium of the fluorescence reaction, FlAsH-EDT$_2$ + peptide ⇆ FlAsH-peptide + 2 EDT depends on the concentration of EDT. Low concentrations of EDT favor the FlAsH-peptide product but also increases the noise due to non-specific binding. Thus two factors, FlAsH-EDT$_2$ and EDT concentrations were considered for the optimization of conditions in the following experiments.

First the detection limit of the tetracysteine peptide and FlAsH system was assessed with 1 μM FlAsH-EDT$_2$. At 10 μM concentration of EDT, a linear increase of the fluorescent signal was observed when the concentration of the peptide increased from 10 nM to 50 nM (Figure 4.9a). However at concentrations lower than 10 nM, it became difficult to distinguish between the concentrations of the peptide based on fluorescence signal (Figure 4.9b). These results show that the detection limit of FlAsH and tetracysteine system is in the low nanomolar range.

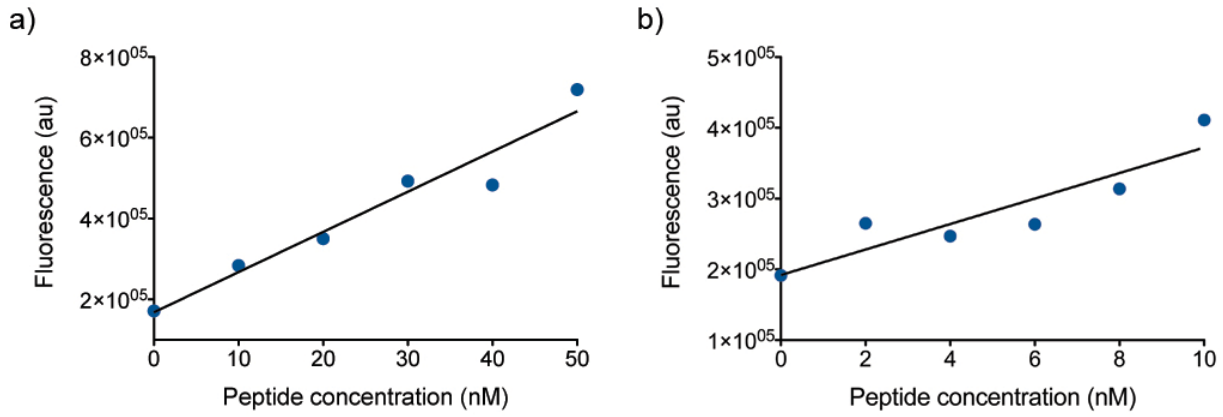



**Figure 4.9** Detection of positive control peptide (FLNCCPGCCMEP). Binding reactions were performed using 1 μM FlAsH-$EDT_2$, in 100 mM Tris-HCl (pH 7.4) containing 3.5 mM TCEP, 10 μM EDT.

*4.2.4 Detection of SMN DNA*

Before evaluating the detection of RNA with the designed system, DNA was evaluated in order to optimize the conditions because DNA is more stable and costs less than RNA. The probes were tested with a 50-mer DNA that contains a sequence encompassing exon 7 and exon 8 (Figure 4.8). To verify that the FlAsH molecule can bind to the designed probes when annealed to the target DNA, I used high concentrations of both probes and target DNA. 1 μM of target DNA was detected using 1 μM of each probes and 25 nM FlAsH-$EDT_2$ and 10 μm EDT. Within 10 minutes, ~9 fold higher fluorescence was observed in the presence of target DNA, FlAsH-$EDT_2$ and the probes compared to the FlAsH only sample. The fluorescence was measured up to 1 hour, and the fluorescence of the solution that contains the target DNA continuously increased. One concern for the experiment was the background fluorescence from the non-specific binding of FlAsH to the probes. Therefore, the signal to noise ratio was calculated as the fluorescence intensity of the solution that contained the target DNA, FlAsH-$EDT_2$, and probes versus the fluorescence intensity of the solution that contained FlAsH-$EDT_2$

and the probes. The signal to noise ratio was highest before 10 minutes and then gradually decreased.

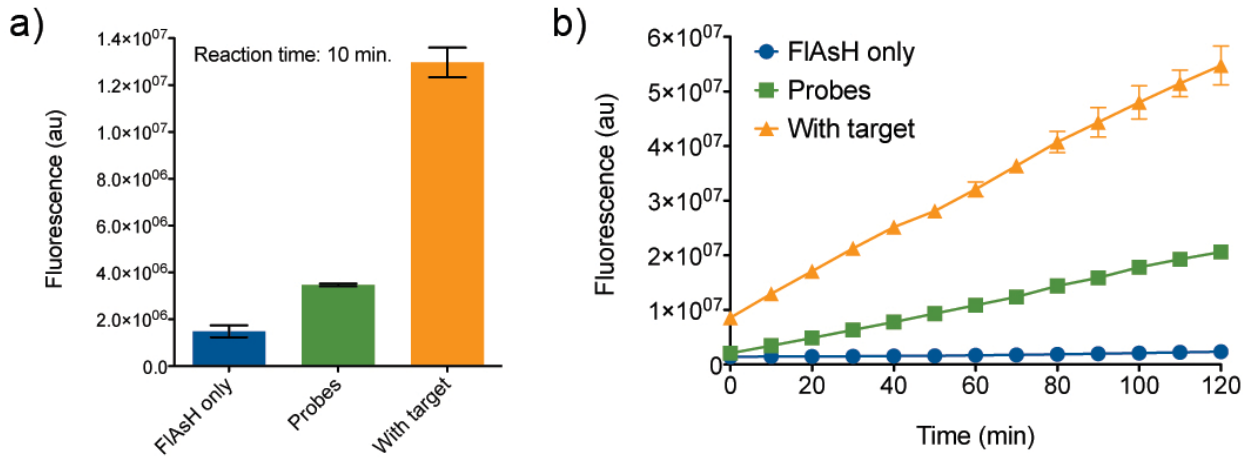



**Figure 4.10** FlAsH fluorescence experiments detecting 1 μM target DNA. a) Fluorescence intensity after 10 minutes of reaction time. b) Kinetics of the reaction measured 2 minutes after mixing and up to 2 hours. "FlAsH only" is a solution containing just the FlAsH-$EDT_2$. "Probes" solution contains FlAsH-$EDT_2$ and both probes. "With target" solution contains FlAsH-$EDT_2$, both probes and the target DNA. Error bars represent SD of triplicates.

Experiments at different concentrations of FlAsH-$EDT_2$ and EDT were performed (Figure 4.11 and 4.12). For all the fluorescence experiments, the target DNA was added first, followed by the probes and then FlAsH-$EDT_2$. 50 nM concentration of target DNA was detected with 50 nM of each probe and different FlAsH concentrations. The signal to noise ratios at 1 μM FlAsH and 100 nM FlAsH are both 2.1. As the FlAsH concentration was decreased further the signal to noise ratio went down and at 1 nM FlAsH concentration the probe signal was higher than the solution that contained the target DNA. 1 μM FlAsH was chosen for further studies because the signal to noise ratio was highest under these conditions and the background fluorescence of the probes compared to the FlAsH only solution was lowest. In contrast to

FlAsH, the concentration of EDT did not have a significant effect on the signal to noise ratio. At four different concentrations of EDT the signal to noise ratio ranged from 2.1 to 2.4 (Figure 4.12).

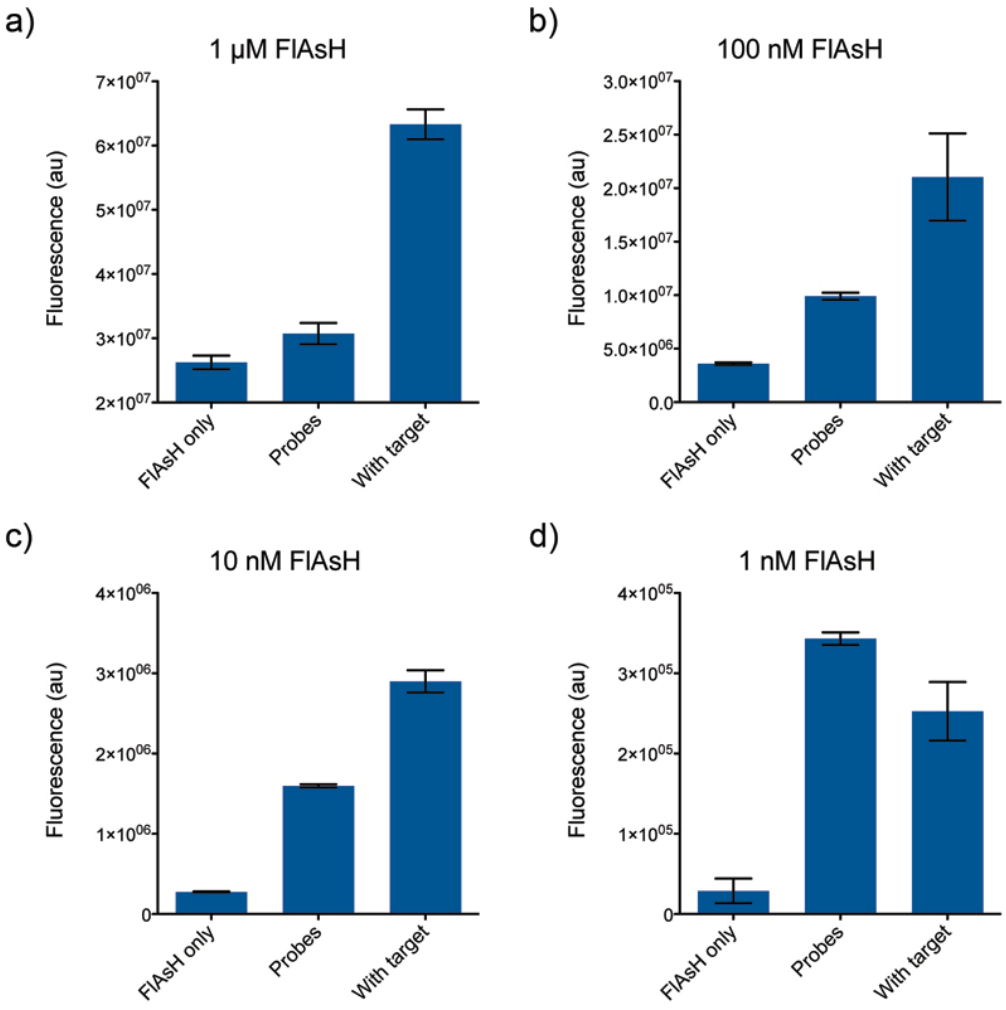



**Figure 4.11** FlAsH fluorescence experiments at different concentrations of FlAsH-$EDT_2$: a) 1 µM, b) 100 nM, c) 10 nM and d) 1 nM. The EDT concentration was 10 µM for all experiments. Error bars represent SD of triplicates.

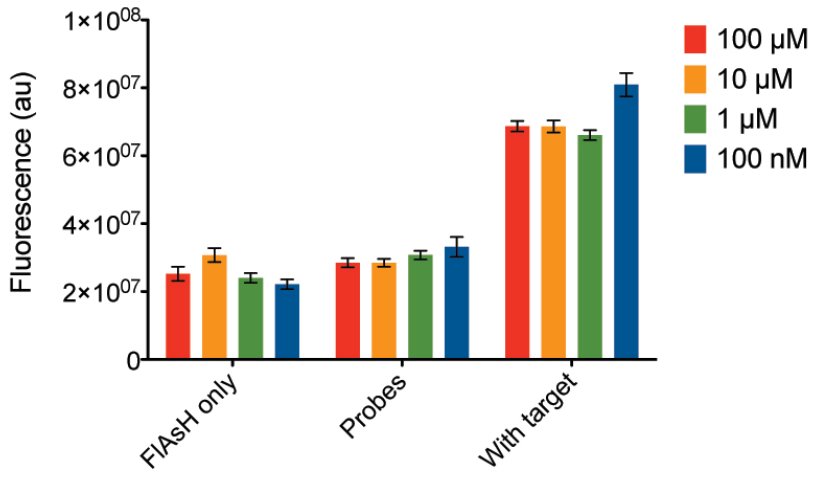



**Figure 4.12** FlAsH fluorescence experiments at different concentrations of EDT. Error bars represent SD of triplicates.

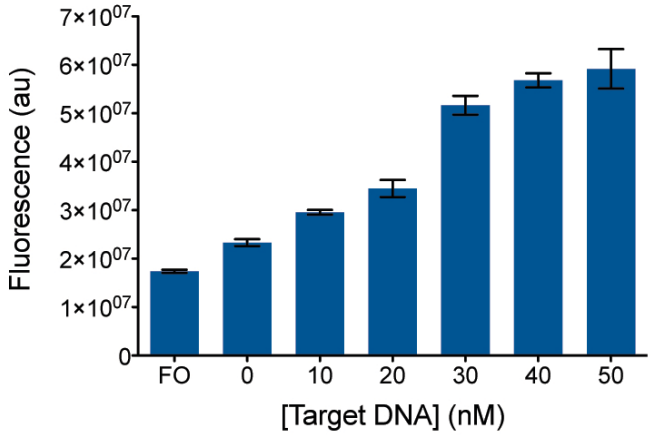



**Figure 4.13** Detection of different concentrations of target DNA. FO is FlAsH only. Error bars represent SD of triplicates.

The ability of these probes to detect different concentrations of the target DNA was assessed. Concentrations ranging from 10 nM to 50 nM with 10 nM increments were detected (Figure 4.13). The fluorescence intensity corresponds to the concentration of the target DNA.

However at 40 nM the fluorescence intensity saturated. Thus, careful adjustment of the concentration of the probes is required to quantitatively detect the target DNA.

*4.2.5 Detection of SMN RNA transcript*

In order to test if the probes and the FlAsH molecule can detect two different RNA splice variants, *SMN* 678 and *SMN* 68 were studied. Plasmid DNAs that contain two different splice products of the *SMN2* gene were transcribed and their sizes confirmed by denaturing gel electrophoresis with RNA size markers. These transcripts were tested with the same buffer conditions used for DNA detection using the 50-mer *SMN* DNA as a comparison. As shown in Figure 4.14, the probes were able to detect the *SMN* 678 transcript within an hour with slightly lower signal to noise ratio than the 50-mer *SMN* DNA. Whereas, the fluorescent intensity of the other splice product, *SMN* 68 did not increase above the background level. A kinetic analysis performed over 13 hours is shown in Figure 4.14a and the corresponding signal to noise ratio is shown in Figure 4.14b. After 13 hours, SMN 678 transcript and 50-mer SMN DNA showed 15-fold and 12-fold increase in fluorescence respectively. The signal to noise ratio reached a maximum at 4 hours and decreased over time. A decrease in the signal to noise ratio over a longer period of time may be due to non-specific binding of the FlAsH molecule to the dicysteine probes and degradation of FlAsH in aqueous solution. The 50-mer *SMN* DNA, compared to *SMN* 678 RNA transcript, shows a faster increase in fluorescence thus resulting in a higher signal to noise ratio up to 7 hours. There are two differences between the DNA detection and the RNA transcript detection. The target DNA forms a DNA-DNA duplex while the target RNA forms a RNA-DNA duplex, and the *SMN* 678 RNA transcript (299 nucleotides) is longer than the short 50-mer DNA target. The RNA transcript may form a secondary structure that

increases the reaction time and reduces the signal to noise ratio. The same experiment was performed at a higher EDT concentration to decrease the background fluorescence due to non-specific binding (Figure 4.15). However, for both 50-mer DNA and *SMN* 678 RNA, the extent of fluorescence decrease due to the higher EDT concentration exceeded the effect of lower background signal. Thus, reduced signal to noise ratios were observed.

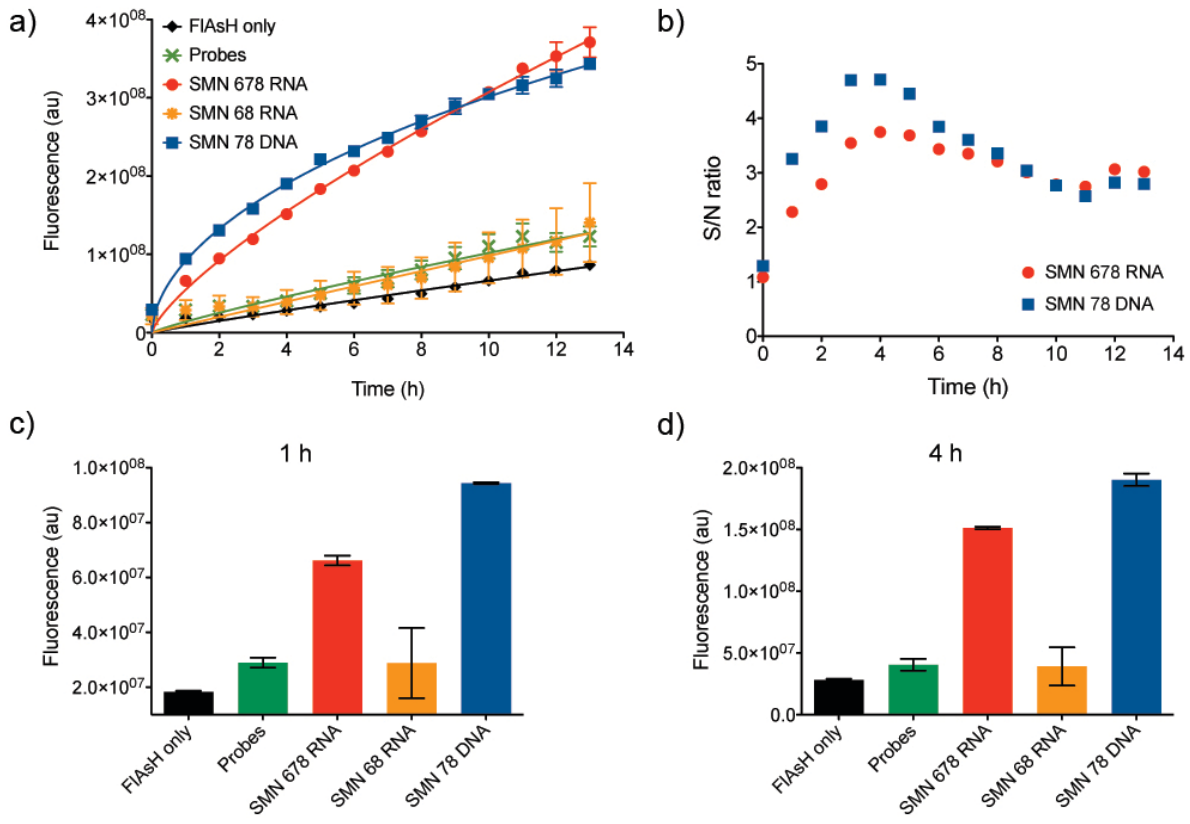



**Figure 4.14** Detection of *SMN* RNA transcripts and 50-mer SMN DNA at 10 μM EDT. a) Kinetic analysis of the detection. b) Signal to noise ratios at different times. Fluorescence intensity of each solution after c) 1 h and d) 4 h of reaction time. Error bars represent SD of triplicates.

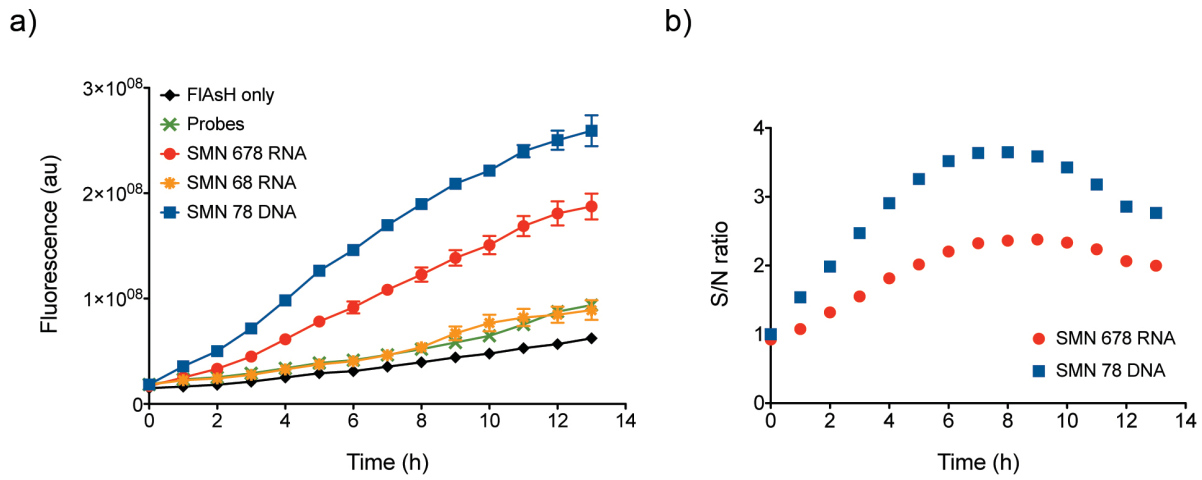



**Figure 4.15** Detection of *SMN* RNA transcripts and 50-mer SMN DNA at 1 mM EDT. a) Kinetic analysis of the detection. b) Signal to noise ratio at different times. Error bars represent SD of triplicates.

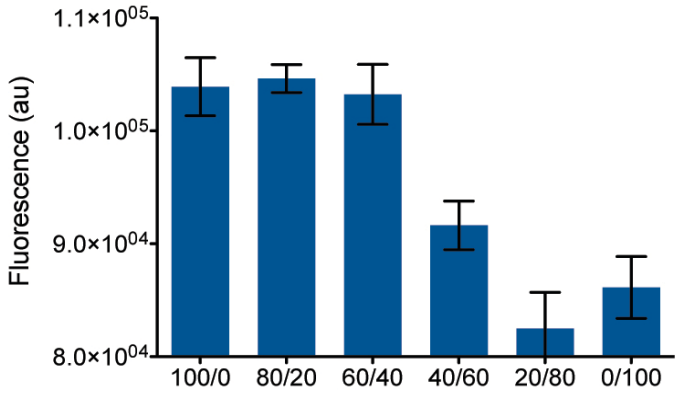



**Figure 4.16** Detection of varied ratios of *SMN* 678 and *SMN* 68 RNA. The P value (student t-test) between 60/40 and 40/60 is less than 0.05. Error bars represent SD of triplicates.

To find small molecules that may increase or decrease an RNA splice variant, the probes need to detect different ratios of the two alternative splice products. *SMN* 678 RNA and *SMN* 68 RNA ratio was varied and the probes and FlAsH-$EDT_2$ were added. As shown in Figure 4.16, the probes were able to distinguish between "high" and "low" ratios but were not able to precisely distinguish each ratio.

*4.2.6 ReAsH Fluorescence Experiments*

ReAsH emits away from the blue and green zone of spectrum in which many cells autofluoresce. Thus, for cell studies, ReAsH has an advantage over FlAsH. ReAsH-$EDT_2$ was synthesized and quality control experiments were performed as previously described by Adams *et al.*[35] Although the ESI-TOF result and NMR confirmed the product compound, a standard 1000-fold increase in fluorescence upon the addition of the control peptide was not observed. Instead a maximum of 152-fold increase in fluorescence was obtained. I proceeded with this ReAsH-$EDT_2$ compound to observe the signal change with the probes and target DNA. An approximately 2-fold increase in fluorescence in the presence of 200 nM target and probes was observed and a slightly higher fluorescence increase with the control peptide (2.6-fold) (Figure 4.17). This indicates that these probes could be used with ReAsH and is not limited to FlAsH. However experiments need to be repeated with higher quality ReAsH-$EDT_2$ to further confirm this preliminary data.

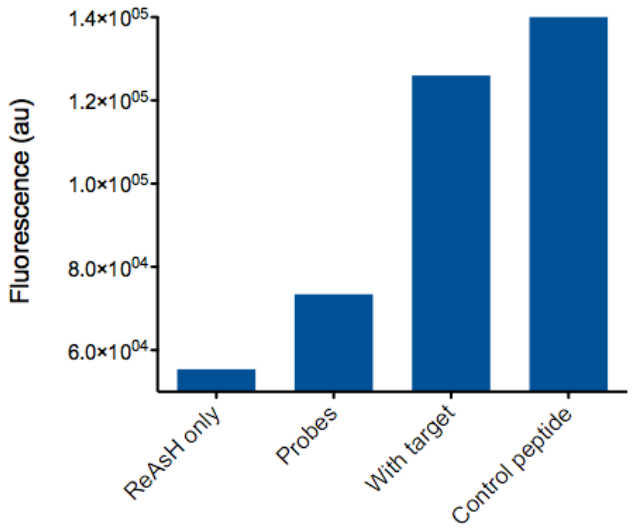



**Figure 4.17** Detection of target DNA with ReAsH. The final concentration of each components were 1μM ReAsH-$EDT_2$, 200 nM target DNA, and 200 nM probes.

### *4.3 Conclusion*

A set of novel nucleic acid detection probes were designed, synthesized and characterized. This method includes the use of biarsenical dye compounds and two antisense oligonucleotides that contain dicysteine moiety at the 3' or 5' end. Fluorescence increase upon addition of target nucleic acid shows that FlAsH compound can efficiently bind to split dicysteine moieties that are brought together by annealing of two antisense oligonucleotides to a target nucleic acid. The probes are capable of detecting micromolar concentration of DNA in less than 10 minutes and as low as nanomolar concentrations of RNA in an hour of reaction time. The signal to noise ratio reached its maximum in 4 hours for both DNA and RNA. Although the probes were not able to accurately distinguish between each concentration, they were able to report "high" or "low" concentration. Careful control of the concentration of the probes may give more quantitative results. Due to the ease of design and synthesis of these probes, they may potentially be used in high-throughput screening assays to identify small molecule modulators of RNA splicing.

### *4.4 Materials and Methods*

#### Materials

All DNAs and RNAs were purchased from Integrated DNA technology (IDT) and were purified by HPLC or PAGE. All solvents were obtained from Fisher Scientific. The trifluoroacetic acid (TFA), mercuric oxide (HgO), 2-(N-morpholino)ethane-sulfonic acid (MES), piperidine, 2,4,6-collidine and ethane dithiol (EDT) were obtained from Sigma-Aldrich. Fmoc-Glu(OtBu)-OH, Fmoc-Met-OH, Fmoc-Cys(StBu)-OH, Fmoc-Gly-OH, Fmoc-Pro-OH, Fmoc-Asn(Trt)-OH, Fmoc-Leu-OH, Fmoc-Phe-OH, Fmoc-Tyr-OH), coupling reagents, and resins were purchased from Chem Impex or Novabiochem.

**Synthesis and quantification of FlAsH-EDT$_2$ and ReAsH-EDT$_2$**

FlAsH-EDT$_2$ and ReAsH-EDT$_2$ were synthesized and purified as previously described.[35] The final product was dissolved in DMSO. The concentration of the DMSO stock solution was determined by diluting the sample into a 0.1 M NaOH solution. The UV absorption was measured at 496 nm for FlAsH-EDT$_2$ and 579 nm for ReAsH-EDT$_2$ and the concentration was calculated using the extinction coefficients reported previously: 69,500 $M^{-1}cm^{-1}$ and 63,000 $M^{-1}cm^{-1}$ respectively. The DMSO stock was aliquoted and stored at -20 °C covered in foil to protect the sample from light.

Sequences of the oligonucleotides used in this study are listed.

Exon 7 probe: 5'-TCCTTAATTTAAGGAATGTG-3'

Exon 8 probe: 5'-CTCTATGCCAGCATTTC-3'

50-mer *SMN* 78 DNA:

5'-GTGCTCACATTCCTTAAATTAAGGAGAAATGCTGGCATAGAGCAGCACTA-3'

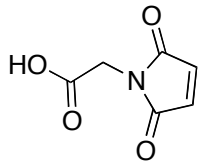
*SMN*2 678 RNA transcript:

CCCGAGAUUCCCCCACCACCUCCCAUAUGUCCAGAUUCUCUUGAUGAUGCUGAUG
CUUUGGGAAGUAUGUUAAUUUCAUGGUACAUGAGUGGCUAUCAUACUGGCUAUU
AUAUGGGUUUCAGACAAAAUCAAAAAGAAGGAAGGUGCUCACAUUCCUUAAAUU
AAGGAGAAAUGCUGGCAUAGAGCAGCACUAAAUGACACCACUAAAGAAACGAUC
AGACAGAUCUGGAAUGUGAAGCGUUAUAGA**C**GAUAACUGGCCU

*SMN*2 68 RNA transcript:

CCCGAGAUUCCCCCACCACCUCCCAUAUGUCCAGAUUCUCUUGAUGAUGCUGAUG
CUUUGGGAAGUAUGUUAAUUUCAUGGUACAUGAGUGGCUAUCAUACUGGCUAUU
AUAUGGAAAUGCUGGCAUAGAGCAGCACUAAAUGACACCACUAAAGAAACGAUC
AGACAGAUCUGAAUGUGAAGCGUUAUAGA**C**GAUAACUGGCCU

**Synthesis of the maleimide linker**



N-glyclylmaleimide was synthesized as previously described.[34]

**Fmoc solid phase peptide synthesis (SPPS)**

Control peptide (YFLNCCPGCCMEP), negative control peptide (GCCMEPY) and the dicysteine peptide (GCC) were manually synthesized using standard Fmoc solid-phase peptide synthesis.[36] Fmoc-amino acid (3 eq to resin loading) and HCTU (3 eq) were dissolved in DMF (4 mL), and to this solution NMM (5 eq) was added and pre-activated for 5 min. The activated amino acid solution was transferred to pre-swelled resin and reacted for 2 h. The completion of couplings was confirmed by a negative Kaiser test. The subsequent removal of Fmoc group was done using 20% piperidine in DMF. Resin, trityl group and OtBu group were removed simultaneously by treatment with a cocktail containing TFA/TIS/$H_2O$ (95:2.5:2.5). For the GCC peptide, the maleimide linker was coupled using the same coupling conditions described above before the cleavage of the resin. All the peptides were purified by reverse phase HPLC using a semi-preparative C18 column (Vydac Protein & peptide C18 (218TP1010), 10 μm, 10 x 250 mm). HPLC method: Flow rate 2 mL/min, detection at 245 nm. Gradient: Starting from 10% MeCN the concentration of MeCN was increased to 80% over 40 min, 80% MeCN was held steady for 60 min. followed by ramping down to 10% MeCN over 10 min. The StBu groups of the control peptide and the negative control peptide were deprotected in a solution of DMF/0.1M ammonium bicarbonate with 20 equiv. of DTT at room temperature overnight. The pH of the reaction mixture was adjusted to 2 with acetic acid and was purified directly by reverse phase

HPLC using the same HPLC method described above. The fractions containing the product were concentrated *in vacuo* and lyophilized. The identities of the positive and negative control peptides were confirmed with low resolution MALDI-TOF and the identity of the dicysteine peptide was confirmed with low resolution ESI-TOF.

Positive control peptide ($M+H^+$): Calculated 1480.8, found 1480.2.

Negative control peptide ($M+H^+$): Calculated 802.9, found 801.5.

GC(StBu)C(StBu)-L peptide ($M+H^+$): Calculated 595.8, found 595.3.

**Synthesis of the probes**

The thiol protecting group of the ordered antisense DNA was removed in a solution of TE buffer (pH 8.0) with 100 mM DTT for 4h or overnight. The deprotected oligonucleotide was purified using amicon® ultra centrifugal filters (MWCO 3000). The reaction mixture was added to the filter and was centrifuged for 40 min. Autoclaved water was added to the filter and centrifuged again for 40 min. This was repeated at least three times. In the last round centrifugation, 0.1 M TTEA buffer was added instead of water for the next reaction. The purified solution of DNA and 10-fold excess of the maleimide linker coupled GCC peptide were allowed to react at room temperature in 1 mL of 20% 0.1 M TEAA/acetonitrile overnight under nitrogen. The resulting product was again purified using amicon® ultra centrifugal filters (MWCO 3000) as decribed above. The last round of purification was performed with TE buffer (pH 8.0). The StBu groups were deprotected in a solution of TE buffer (pH 8.0) with 100 mM DTT overnight. The deprotected StBu groups and excess of DTT were removed with amicon® ultra centrifugal filters (MWCO 3000). The concentration of the final product was determined

using UV spectroscopty and the extinction coefficient of the DNA was provided by the manufacturer. The probes were confirmed with ESI-TOF.

**Transcription of SMN 68 and SMN 678**

Plasmid DNAs that encode SMN2 exon 678 and SMN2 exon 68 were each ordered from IDT. The sequences of the two pDNAs were confirmed at University of illinois at Urbana-Champaign sequencing center. SMN plasmid DNAs were linearized for transcription using restriction enzyme, Not I (Invitrogen) in a total of 20 μL buffered solution provided by the manufacturer. The reaction was left overnight at 37 °C for complete cleavage. The cleaved products were purified by running an agarose gel and extracting the bands using a gel extraction kit (Qiagen). The linearized plasmid DNAs were transcribed following the MEGAScript (Ambion) instructions. The transcripts were confirmed using a 5% denaturing polyacrylamide gel (8.3 × 7.3 cm). The gel was pre-ran at 100 V and the samples and RNA markers (RNA Century$^{TM}$ Plus Markers, Ambion) were ran at 80 V for 90 min.

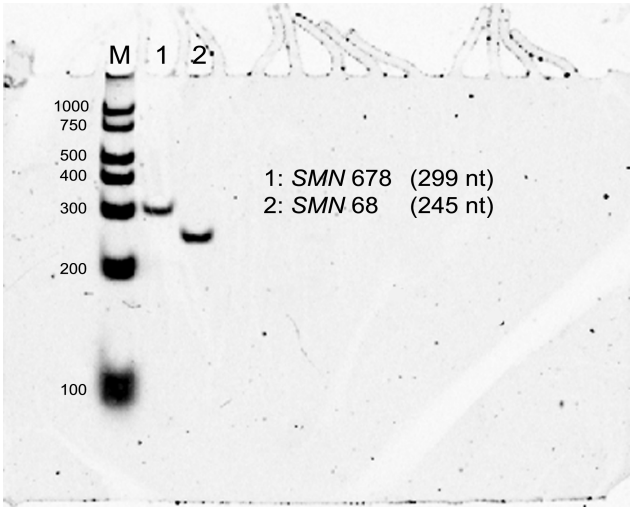



**Figure 4.18** Denaturing PAGE of *SMN* 678 and *SMN* 68 RNA transcripts. M is marker, lane 1 is *SMN* 678, and lane 2 is *SMN* 68. The numbers on the left side of the RNA marker lane are the number of nucleotides.

**General procedure for fluorescence experiments**

TTEM buffer (100 mM Tris-HCl, 3.5 mM TCEP, 10 μM EDT, and 10 mM $MgCl_2$) was freshly prepared for each experiment from the following stock solutions: 1M Tris-HCl (pH 7.4), 0.5 M TCEP, 10 mM EDT, 1M $MgCl_2$. The 10 mM EDT solution was prepared fresh for each experiment by adding 0.84 μL of EDT into 1 mL of dry DMSO. Typically 2~4 mL of TTEM buffer was made and was degassed by vacuum sonication for 20 min. In a 384 well plate, the buffer was added followed by the probes and the target nucleotide. FlAsH-$EDT_2$ was added last and the solution was left at room temperature covered with an aluminum foil for indicated amount of time. For longer incubation times (> 2h), the plates were sealed to prevent evaporation of the solution.

### *4.5 References*


1 Matlin, A. J.; Clark, F.; Smith, C. W. Understanding alternative splicing: towards a cellular code, *Nat Rev Mol Cell Biol* **2005**, *6*, 386.

2 Sultan, M.; Schulz, M. H.; Richard, H.; Magen, A.; Klingenhoff, A.; Scherf, M.; Seifert, M.; Borodina, T.; Soldatov, A.; Parkhomchuk, D.; Schmidt, D.; O'Keeffe, S.; Haas, S.; Vingron, M.; Lehrach, H.; Yaspo, M. L. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome, *Science* **2008**, *321*, 956.

3 Buratti, E.; Baralle, F. E. Influence of RNA secondary structure on the pre-mRNA splicing process, *Mol Cell Biol* **2004**, *24*, 10505.

4 Kornblihtt, A. R. Chromatin, transcript elongation and alternative splicing, *Nat Struct Mol Biol* **2006**, *13*, 5.

5 Wang, G. S.; Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery, *Nat Rev Genet* **2007**, *8*, 749.

6 Caceres, J. F.; Kornblihtt, A. R. Alternative splicing: multiple control mechanisms and involvement in human disease, *Trends Genet* **2002**, *18*, 186.

7 Hicks, M. J.; Lam, B. J.; Hertel, K. J. Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays, *Methods* **2005**, *37*, 306.

8 Kaida, D.; Motoyoshi, H.; Tashiro, E.; Nojima, T.; Hagiwara, M.; Ishigami, K.; Watanabe, H.; Kitahara, T.; Yoshida, T.; Nakajima, H.; Tani, T.; Horinouchi, S.; Yoshida, M. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA, *Nat Chem Biol* **2007**, *3*, 576.

9 Kaida, D.; Berg, M. G.; Younis, I.; Kasim, M.; Singh, L. N.; Wan, L.; Dreyfuss, G. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation, *Nature* **2010**, *468*, 664.

10 Zhang, M. L.; Lorson, C. L.; Androphy, E. J.; Zhou, J. An in vivo reporter system for measuring increased inclusion of exon 7 in SMN2 mRNA: potential therapy of SMA, *Gene Ther* **2001**, *8*, 1532.

11 Chang, J. G.; Hsieh-Li, H. M.; Jong, Y. J.; Wang, N. M.; Tsai, C. H.; Li, H. Treatment of spinal muscular atrophy by sodium butyrate, *Proc Natl Acad Sci U S A* **2001**, *98*, 9808.

12 Andreassi, C.; Jarecki, J.; Zhou, J.; Coovert, D. D.; Monani, U. R.; Chen, X.; Whitney, M.; Pollok, B.; Zhang, M.; Androphy, E.; Burghes, A. H. Aclarubicin treatment restores SMN levels to cells derived from type I spinal muscular atrophy patients, *Hum Mol Genet* **2001**, *10*, 2841.

13 Soret, J.; Bakkour, N.; Maire, S.; Durand, S.; Zekri, L.; Gabut, M.; Fic, W.; Divita, G.; Rivalle, C.; Dauzonne, D.; Nguyen, C. H.; Jeanteur, P.; Tazi, J. Selective modification of alternative splicing by indole derivatives that target serine-arginine-rich protein splicing factors, *Proc Natl Acad Sci U S A* **2005**, *102*, 8764.

14 Nasim, M. T.; Eperon, I. C. A double-reporter splicing assay for determining splicing efficiency in mammalian cells, *Nat Protoc* **2006**, *1*, 1022.

15 Stoilov, P.; Lin, C. H.; Damoiseaux, R.; Nikolic, J.; Black, D. L. A high-throughout screening strategy identifies cardiotonic steroids as alternative splicing modulators, *Proc Natl Acad Sci U S A* **2008**, *105*, 11218.

16 Wang, K.; Tang, Z.; Yang, C. J.; Kim, Y.; Fang, X.; Li, W.; Wu, Y.; Medley, C. D.; Cao, Z.; Li, J.; Colon, P.; Lin, H.; Tan, W. Molecular engineering of DNA: molecular beacons, *Angew Chem Int Ed Engl* **2009**, *48*, 856.

17 Xu, Y.; Karalkar, N. B.; Kool, E. T. Nonenzymatic autoligation in direct three-color detection of RNA and DNA point mutations, *Nat Biotechnol* **2001**, *19*, 148.

18 Silverman, A. P.; Abe, H.; Kool, E. T. Quenched autoligation probes, *Methods Mol Biol* **2008**, *429*, 161.

19 Franzini, R. M.; Kool, E. T. Efficient nucleic acid detection by templated reductive quencher release, *J Am Chem Soc* **2009**, *131*, 16021.

20 Nakayama, S.; Yan, L.; Sintim, H. O. Junction probes - sequence specific detection of nucleic acids via template enhanced hybridization processes, *J Am Chem Soc* **2008**, *130*, 12560.

21 Yan, L.; Nakayama, S.; Yitbarek, S.; Greenfield, I.; Sintim, H. O. Isothermal detection of RNA with restriction endonucleases, *Chem Commun* **2011**, *47*, 200.

22 Griffin, B. A.; Adams, S. R.; Tsien, R. Y. Specific covalent labeling of recombinant protein molecules inside live cells, *Science* **1998**, *281*, 269.

23 Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications, *J Am Chem Soc* **2002**, *124*, 6063.

24 Gaietta, G. M.; Giepmans, B. N.; Deerinck, T. J.; Smith, W. B.; Ngan, L.; Llopis, J.; Adams, S. R.; Tsien, R. Y.; Ellisman, M. H. Golgi twins in late mitosis revealed by genetically encoded tags for live cell imaging and correlated electron microscopy, *Proc Natl Acad Sci U S A* **2006**, *103*, 17777.

25 Nakanishi, J.; Takarada, T.; Yunoki, S.; Kikuchi, Y.; Maeda, M. FRET-based monitoring of conformational change of the beta2 adrenergic receptor in living cells, *Biochem Biophys Res Commun* **2006**, *343*, 1191.

26 Luedtke, N. W.; Dexter, R. J.; Fried, D. B.; Schepartz, A. Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH, *Nat Chem Biol* **2007**, *3*, 779.

27 Krishnan, B.; Gierasch, L. M. Cross-strand split tetra-Cys motifs as structure sensors in a beta-sheet protein, *Chem Biol* **2008**, *15*, 1104.

28 Burghes, A. H.; Beattie, C. E. Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick?, *Nat Rev Neurosci* **2009**, *10*, 597.

29 Frugier, T.; Nicole, S.; Cifuentes-Diaz, C.; Melki, J. The molecular bases of spinal muscular atrophy, *Curr Opin Genet Dev* **2002**, *12*, 294.

30 Clermont, O.; Burlet, P.; Lefebvre, S.; Burglen, L.; Munnich, A.; Melki, J. SMN gene deletions in adult-onset spinal muscular atrophy, *Lancet* **1995**, *346*, 1712.

31 Cartegni, L.; Krainer, A. R. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1, *Nat Genet* **2002**, *30*, 377.

32 Burglen, L.; Spiegel, R.; Ignatius, J.; Cobben, J. M.; Landrieu, P.; Lefebvre, S.; Munnich, A.; Melki, J. SMN gene deletion in variant of infantile spinal muscular atrophy, *Lancet* **1995**, *346*, 316.

33 Sendtner, M. Therapy development in spinal muscular atrophy, *Nat Neurosci* **2010**, *13*, 795.

34 Brownlee, J. M.; Carlson, E.; Milne, A. C.; Pape, E.; Harrison, D. H. Structural and thermodynamic studies of simple aldose reductase-inhibitor complexes, *Bioorg Chem* **2006**, *34*, 424.

35 Adams, S. R.; Tsien, R. Y. Preparation of the membrane-permeant biarsenicals FlAsH-EDT2 and ReAsH-EDT2 for fluorescent labeling of tetracysteine-tagged proteins, *Nat Protoc* **2008**, *3*, 1527.

36 Hood, C. A.; Fuentes, G.; Patel, H.; Page, K.; Menakuru, M.; Park, J. H. Fast conventional Fmoc solid-phase peptide synthesis with HCTU, *J Pept Sci* **2008**, *14*, 97.